



A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-709 <TIGR>  
 A:Cross-references: GB:U32710; GB:L42023; NID:g1573200; PID:g1573211; TIGR:HI0248

alignment\_scores:  
 Quality: 2494.00 Length: 701  
 Ratio: 4.075 Gaps: 5  
 Percent Similarity: 87.304 Percent Identity: 67.618

alignment\_block:  
 US-09-303-518D-653 x C64057 ..

Align seg 1/1 to: C64057 from: 1 to: 709

```

79 TACTTAGCATATGCTCGTTCGGCATCTGCCCAAGCCGCGGG 128
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
10 PheLeuThrAlaCysValSerLeuGlyLeuAlaSerGlnAlaTrpAlaG1 26
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
129 ACACACTTATTTCCGCATCACTACCACTACTATCGCGACTTCCCGAAA 178
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
26 yHisThrTyPheGlyIleAspTyGlnTyTrArgAspPheAlaGlu 43
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
179 ATAAAGGCAAGTTTCAGTCGGGGGGAAGATATTGAGTTTACAACAAA 228
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
43 snLysGlyLysPheThrValGlyAlaLysAsnIleGluValTyAsnLys 59
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
229 AAAGGGAGTTGGTCGGAATTCGATGACGAAAGCCCGATGATGATT 278
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
60 GluGlyGlnLeuValGlyThrSerMetThrLysAlaProMetIleAspPh 76
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
279 TTCTGTGGTATCGCTAAGCGCGTGGCGCATTTGGCGGCATCAATATA 328
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
76 eSerValSerArgAsnGlyValAlaAlaLeuValGlyAspGlnTyri 93
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
329 TTGTGAGCGTGGCACATAACGGCGGTATACAAATGTTGTTGGTGG 378
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
93 leValSerValAlaHisAsnGlyGlyTyAsnAspValAspPheGlyAla 109
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
379 GAGGGAAGCAATCCCGATCAGCACCGCTTTCTTACCAATTTGCAAG 428
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
110 GluGlyArgAsnProAspGlnHisArgPheThrTyGlnIleValLysAr 126
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
429 AATATATTATAAGCAGGACCTAACGGCCATCTTATGGCGCGATTATC 478
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
126 gAsnAsnTyGlnAlaTrpGluArgLysHisProTyArgAspGlyAspTyH 143
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
479 ATATGCGCGCTTTGCACAAATTTGTACAGATGCGACACCTCTTGAGATG 528
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
143 IsMetProArgLeuHisLysPheValThrGluAlaGluProValGlyMet 159
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
529 ACCAGTTATATGATGGTGGGAATACGCTGATTTAAATATAATACCTGA 578
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
160 ThrThrAsnMetAspGlyLysValTyAlaAspArgGluAsnTyProG1 176
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
579 TCGTGTTCGAATCGGACGACGACAGCAATATTGGCGGTCTGATGAAGAG 628
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
176 uArgValArgIleGlySerGlyArgGlnTyTrArgThrAspLysAspG 193
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
629 AACCCATATACCGGAAAGTTTCATATCATATTCGAAGCGCATTTCTTGG 678
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
193 luGluThrAsnValHisSerSerTyTyTyValSerGlyAlaTyArgTy 209
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
679 CTCGTGCGTGGCAATACCTTTGCACAAAATGGAATGAGTGGTGGCAGAC 728
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
210 LeuThrAlaGlyAsnThrHisThrGlnSerGlyAsnGlyAsnGlyThrVa 226
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
729 CAACCTTAGTAGCGAAATAATTAACATAGCCCATATGTTTTCACAA 778
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
226 lAsnLeuSerGlyAsnValValSerProAsnHisTyTyGlyProLeuProT 243
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
779 CAGGAGGCTCATTTGGCGCAGTGGCTCACCACCAATGTTTATCTATGATGCC 828

```

```

243 hrGlyGlySerLysGlyAspSerGlySerProMetPheIleTyAspAla 259
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
829 CAAAGCAAAAGTGGTAAATTAATGGGGTATTGCAACAGGCAACCCCTA 878
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
260 LysLysLysGlnTrpLeuIleAsnAlaValLeuGlnThrGlyHisProPh 276
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
879 TATAGGAAAAGCAATGGCTTCCAGCTAGTTCGTAAGATGTTGTTCTATG 928
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
276 ePheGlyArgGlyAsnGlyPheGlnLeuIleArgGluGluTrpPheTyra 293
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
929 ATGAAATCTTTCGTCGGAGATACCCATTTCAGTATTC.....TACGAACCA 972
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
293 snGluValLeuAlaValAspThrProSerValPheGlnArgTyriIlePro 309
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
973 CATCAAAATGGAAATACTTTTAAACGACAAATAATAATGGCGCAGGAAA 1022
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
310 ProIleAsnGlyHisTySerPheValSerAsnAsnAspGlyThrGlyLy 326
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1023 AATCGAT...GCCAAACATAAACAATACTTCTTACCTTATAGATTAAAAA 1069
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
326 sLeuThrLeuThrArgProSerLysAspGlySerLysAlaLysSerGluV 343
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1070 CACGAAACGTTCAATTGTTTAACTTCTTATCCGAGACAGCAAGAGAA 1119
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
343 aGlyThrValLysLeuPheAsnProSerLeuAsnGlnThrAlaLysGlu 359
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1120 CCGTGTATCATGCTGTCAGGTGGGTCAACAGTTTATCGACCCAGCTGAA 1169
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
360 HisVal...LysAlaAlaAlaGlyTyAsnIleTyGlnProArgMetG1 375
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1170 TAATGGAGAAATATTTCTTATTGACAAAGGAAAGGTCGAATTGATAC 1219
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
375 uTyrgLysAsnIleTyLeuGlyAspGlnGlyLysGlyThrLeuThrI 392
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1220 TTACGACAAACATCAACGAGCGGGCGGCTTGTATTATTTGAGGGTAAT 1269
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
392 lGluLysAsnIleAsnGlnGlyAlaGlyGlyLeuTyPheGluGlyAsn 408
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1270 TTTACGGTCTCGCCTTAAA...AACAAACAAACGTCGGCAAGCGCGGGGT 1316
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
409 PheValValLysGlyLysGlnAsnAsnIleThrTrpGlnGlyAlaGlyVa 425
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1317 TCATATCATGTGATCGCAGTACGCTTGTGAAAGTAAACGGCGTCGCAA 1366
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
425 lSerIleGlyGlnAspAlaThrValLglutrpLysValHisAsnProGluA 442
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1367 ACACGCGCTGTCCAAAATCGGCAAGGACGCTGCTGTTCAAGCCAAA 1416
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
442 snAspArgLeuSerLysIleGlyIleGlyThrLeuLeuValAsnGlyLys 458
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1417 GGGGAAAACCAAGCTCGTCAAGCTGGGCGGCGGTAAGTCACTCTTAGA 1466
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
459 GlyLysAsnLeuGlySerLeuSerAlaGlyAsnGlyLysValIleLeuAs 475
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1467 TCAGCAGCGGCGAGTCAAGCAAGCAAAACAAAGCCCTTTAGTGAATTCGGCT 1516
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
475 pGlnGlnAlaAspGluAlaGlyGlnLysGlnAlaPheLysGluValGlyI 492
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1517 TGTGTCAGCGCAGGGGAGCGGTGCAACTGAATGCCGATTAATCAGTTCAAC 1566
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
492 leValSerGlyArgAlaThrValGlnLeuAsnSerThrAspGlnValAsp 508
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1567 CCGCACAACACTCTATTTTCGCTTTCGCGCGGACGCTTTGGATTTCGAACGG 1616
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
509 ProAsnAsnIleTyPheGlyPheArgGlyGlyArgLeuAspLeuAsnG1 525
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1617 GCATTTCGCTTTCGTCACCGCATTCAAAATACCGATGAAGGGCGGATGA 1666
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
525 yHisSerLeuThrPheLysArgIleGlnAsnThrAspGluGlyAlaMetI 542
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1667 TGTGTAACCACTCAAGCAAGAAATCCACGCTTACCATTTACAGCAAT 1716
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

542 leValAsnHisAsnThrThrGlnValAlaAsnIleThrIleThrGlyAsn 558

1717 AAAGATATTACTACACCGGCAATAAC.....AACAACTTGGATAG 1757

559 GluSerIleThrAlaProSerAsnLysLysAsnIleAsnLysLeuAspTy 575

1758 CAAAAAGAAATTGCCTACACCGTTGGTTGGCGAGAAAGATGCAACCA 1807

575 rSerLysGluIleAlaIleAsnGlyTrpPheGlyGluThrAspLysAsnL 592

1808 AACGAACACGGCGCTCAATCTCAATTACCAACCGGAAGCGGATCGC 1857

592 yHisAsnGlyArgLeuAsnLeuIleTyLysProThrThrGluAspArg 608

1858 ACTTTACTGCTTCCGGCGGAACAAATTTAAACGGCAATATCACGCAAA 1907

609 ThrLeuLeuLeuSerGlyGlyThrAsnLeuLysGlyAspIleThrGlnTh 625

1908 AAACGGCAAACTGTTTTTCAGGGGACAGCCGACCGCACGCGCTACAATC 1957

625 rLysGlyLysLeuPheSerGlyArgProThrProHisAlaIleAsnH 642

1958 ATTTAGGAACGGGTGGTCAAAATGGAAGGTATCCCAAGGAGAAATTC 2007

642 iLeuAspLysArgTrpSerGluMetGluGlyIleProGlnGlyGluIle 658

2008 GTGTGGGCAACAGATTGGATCCACCGCACATTTAAACGGAAACTTCCA 2057

659 ValTrpAspTyAspTrpIleAsnArgThrPheLysAlaGluAsnPheG 675

2058 TATTCAGGCGGCACAAAGCGTGGTTCCCGCAATGTGCCAAAGTGGAG 2107

675 nIleLysGlyGlySerAlaValIleSerArgAsnValSerSerIleGluG 692

2108 GCGATTGGCATTTAGCAATCATCGCCCAAGCAGTTTCGGTGTGCGACGC 2157

692 lYAsnTrpThrValSerAsnAsnAlaAsnAlaThrPheGlyValValPro 708

2158 CAT 2160

709 Asn 709

seq\_name: pir2:A37023

seq\_documentation\_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae

N;Alternate names: immunoglobulin A1 proteinase type 1

C;Species: Haemophilus influenzae

C;Date: 31-Jan-1992 #sequence\_revision 12-Jun-1992 #text\_change 08-Dec-2000

C;Accession: A37023

R;Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Kilian, M.

A;Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemophilus influenzae

A;Reference number: A37023; M01D:89379374

A;Accession: A37023

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1541 <POD>

A;Cross-references: GB:X64357; NID:G43560; PIDN:CAA45708.1; PID:G43561

A;Experimental source: serotype b

C;Superfamily: IgA-specific metalloendopeptidase

C;Keywords: hydrolase; metalloproteinase

64 CGCTTCTCGCCGCTTACTTAGCCATATACCTGTGCTCGGTCCGATCTGC 113  
 5 LysPheLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21  
 114 CCAAGCCGGCGGGACACACITATTTCGGCATCACTACCAATPACTATC 163  
 21 tYrThrGluAlaLeuValArgAspAspValAspTyrGlnIlePheA 38  
 164 CGCACTTTGCCAAAATAAGCAAGTTTGACGTGGGCGGCAAGATATT 213  
 38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54  
 214 GAGGTTCACACAAAAGGGGACGTGTGTCGCAATCGATCAGCAAGC 263  
 55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnG 71  
 264 C...CCGATGATGATTTCTGTGGTATCG...CGTAACGGCGTGGCGG 307  
 71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAla 88  
 308 CATTGGCGGGCGCATCAATATATTGACGGTGGCACAT...AACGGCGG 354  
 88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104  
 355 TATAACAATGTTGATTGTTGGT.....CGGAGGGGAAGCAA 389  
 105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121  
 390 TCCCGATCAGACCGC.....TTTTCTTACCAAAATG 421  
 121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138  
 422 TGAAGAAGAAATAATTATAACAGGAGGACTAAGGC..... 456  
 138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThrThr 154  
 457 .....CATCCTTATGGCGCGGATTCATCATATCGCGCGTTGCA 494  
 155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171  
 495 CAATTTTGTACAGATCGAGAACCTGTTGAG...ATGACCAAGTTATATGG 541  
 171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188  
 542 ATGGGTGAAATACGCTGATTAAATAAATACCCTGATCGTGTTCGAATC 591  
 188 sPAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204  
 592 GGACGAGCGACAGCAATATTGCGGCTCTGATCAACACGAA..... 630  
 205 GlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuIl 221  
 631 .CCCAATAAACCGCAAGTTCATATCAT.....ATTGCAACGC 667  
 221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238  
 668 CATATTCTTGCTCGTGGTGGCAATACCTTTGCACAAAATGGATCAGGT 717  
 238 laTyrThrTygIleAlaGlyThrProTyrLysValAsnHisGluAsn 254  
 718 GGTGGCAGACGTCAACTTAGGTAGGGAATAATTAACATAGC...CCATA 764  
 255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271  
 765 TGGTTTTTACCACAGGA.....GGCTCATTTTGGCGACA 799  
 271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAsps 288  
 800 GTGGCTCACCAANTGTTTATCTATCATGCCCAAAAGCAAAAGTGGTTAAT 849  
 288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTyrLeuPhe 304

850 AATGGGTTATTGCAACAGGCAACCCCTATATAGGAAAAAGCAATGCTT 899  
||| : : : |||  
305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysSer..... 318  
900 CCAGCTAGTTCGTAAGATTGG...TTCTATGATGAAATCTTTGCTGGAG 946  
: : : : : ||| : : : : : ||| : : : : :  
319 .....TrpGlnGluTrpAsnIleTyrLysSerGlnPheThrLysA 332  
947 ATACCATTCAGTATTCTACGACACATCAAAATGGGAAATACATTTTT 996  
||| : : :  
332 spValLeu..... 334  
997 AACGACATATAATGCGCAGGAAATATGATGCAACACATAAACACTA 1046  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
335 .....AsnLysAspSerAlaGlySerLeuIleGlySerLysThrAspTy 349  
1047 TTCTCTACCTATATGATTAATAAACACGACCGCTT.....CAAT 1084  
||| : : : ||| : : : ||| : : :  
349 rSerTrpSerAsnGlyLysThrSerThrIleThrGlyGlyGluLys 366  
1085 TGTTTAATGTTCTTTATCCGAGACAGCAAGAACCTGTTTATCATGCT 1134  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
366 erLeuAsnValAspLeuAlaAspGlyLysAspLysPro..... 378  
1135 GCAGGTGGGTCAACAGTTATCGACCCAGACTGAATATGAGAAATAT 1184  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
379 .....AsnHisGlyLysSerVa 384  
1185 TTCCTTTATTGCAAAAGGAAAGTGAATGATCTTACCAGCAACATCA 1234  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
384 lThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsnIleA 399  
1235 ACCAGGCGCGCGCTTGTATTGAGGTGAATTTTACGGTC...TCG 1281  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
399 spGlnGlyAlaGlyGlyLeuPheGluGlyAspTyrGluValLysGly 415  
1282 CCAAAAACAAAGCAACGTGGCAAGCGCGGCTTCATATCAGTGATGG 1331  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
416 ThrSerAspAsnThrTrpLysGlyAlaGlyValSerValAlaGluG 432  
1332 CAGTACCGTTACTGGAAGTAAACGGCGTGGCAACGACCGCCCTGCA 1381  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
432 yLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeuAla 449  
1382 AATCGGCAAGCGCGCTGCTGTTCAAGCCAAAGGGGAAACCAAGGC 1431  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
449 ySileGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLysGly 465  
1432 TCGGTGAGCGTGGGACCGGTAAAGTCATCTTAGATCAGAGCGGAGCA 1481  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnG 482  
1482 TCAGGCAAAAACAGCCCTTTAGTGAATCGCTTGGTCAGCGCGCAGG 1531  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
482 ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyArgS 498  
1532 GGAGGTGCAACTGAATCGCGGTAATCAGTTCAACCCCGCACAACTCTAT 1581  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
498 erThrLeuValLeuAsnAspLysGlnValAspProAsnSerIleTyr 514  
1582 TTCGGTTTCGCGCGGAGCTTTGATTTGAACGGGCATTCCTTCGTT 1631  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
515 PheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531  
1632 CCACCCCATTAATACCGATGCAAGGCGGATGATGTCACACCAATC 1681  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
531 eAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsn 548  
1682 AAGCAAAAGATCCCGTTTACCATACAGGCAATAAAGATATTACT... 1728  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
548 etThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThrAsp 564  
1728 ..... 1728

565 ProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnPr 581  
1728 ..... 1728  
581 oTyrAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuG 598  
1728 ..... 1728  
598 luAsnTyrThrTyrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614  
1729 .....ACAACGGCAATAAC..... 1743  
615 LeuProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGlyLy 631  
1744 .....AACRACTTGGATAGCAAAA 1762  
631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluA 648  
1763 AAGAAATTCGTACAAACGGTTGGTGGGAGAAAGATGCAACCAAAACG 1812  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
648 rgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGly...LysAsn 663  
1813 AACGGCGGCTCAATCTGAATTACCAACCGGAGAGACCGGATCGCACTTT 1862  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
664 AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPh 680  
1863 ACTGCTTTCGGCGGAAACAAATTTAAACGCAATATACGCAACAAACG 1912  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
680 eleuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluLysG 697  
1913 GCAACTGTTTTTTCAGCGGACGACCGCACCGCCCTACATCATTTA 1962  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
697 lyThrLeuPheLeuSerGlyArgProHisAlaArgAspIleAla 713  
1963 GGAACGGGTGGTCAAAATGGAAGT.....ATCCCAAGGAGAAAT 2006  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
714 GlyLeuSerSerThrLysLysAspProHisPheAlaGluAsnAsnGluVa 730  
2007 CGTGTGGGACACGATTGGATCGACCGCACATTTAAAGCGGAAACTTCC 2056  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
730 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrMeta 747  
2057 ATATTACGGCGGACACGCGGTGTTCC...CGCAATGTCGCAAGTG 2103  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
747 snValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnIle 763  
2104 GAAGCGATTGGCATTTAAGCAATCACGCCCAAGCAGTTTTCGTGTCGC 2153  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTyr... 779  
2154 ACCGATCAAGCCACACAATCTGTACACGTTGCGGACTGGACGGTCTGA 2203  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
780 ....LysThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 795  
2204 CAAGTTGTACGAAAAAACCATCACCGACCATTAAGTGATTGCTTCATTG 2253  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
795 alThrCysThrThrAspLysSer...AspLysAlaLeuAsnSerPhe 810  
2254 AGCAAGACCGACATCAGAGCAATGTGCGCTTGGCCATCAGCGTCAATT 2303  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
811 AsnProThrAsnLeuArgLysAsnValAsnLeuThrGluSerAlaAsnPh 827  
2304 AAATTCACAGGACTTGCACACTCAACGGCAATCTTAGTCAGGCGGAG 2353  
||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :  
827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843  
2354 ACAGCGCTATACGTTACGGCAACGCCCAACCAACCAACCAACCTCAGC 2403  
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
843 snSerGlnValArgLeuThrGluAsn..... 851  
2404 CTCGTGGGCAATGCCAAGCAACATTTAATCAAGCCACATTAACACGCAA 2453



851 ..... 851  
2454 CACATCGCGCTTCGGACAAATGCTTCATTAAATCTAAGCAACAACCGCGTAC 2503  
851 ..... 851  
2504 AAAACGGCAGTCTGACGCTTTCGCACACGCTAAGGCAACGTAAGCCAT 2553  
851 ..... 851  
2554 TCCGCACTCAACGCAATGCTCCCTAGCCGATAAGCAGTATTCCATTT 2603  
851 ..... 851  
2604 TGAACACAGCGCTTTACCGGAAAAATACAGCGCGGCAAGATACGGCAT 2653  
851 ..... 851  
2654 TACACTTAAAGACAGCAATGGAGCTGCGCGGCACGGAATTAGGC 2703  
852 ..... 852  
2704 AATTAAACCTTGACACGCGCATTACACTCAATTCGCCCTATCGACA 2753  
864 GlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAsnSe 880  
2754 CGATCGCGCAGCGCGCAACCGCGCAGTGGCGGAGATGCCCGCGCGGCC 2803  
880 rAsn..... 881  
2804 GTTCGCGCGCTTCCTATTATCCGCTTACGCGGCAACTTCGGCAGATCC 2853  
882 ..... 884  
2854 CGTTTCAACAGCTGACGGTAACGGCAATTAACGGTCAGGGAACATT 2903  
885 LysTyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPh 900  
2904 CCGCTTTATGCGAACTCTTCGGCTACCGCAGCGGCAATTAAGACTGG 2953  
900 eTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspValValVal 917  
2954 CGGAACCTCCGAGGACATTACACTTGGCTGTCAACAATACCGGCAAC 3003  
917 hrLysSerAlaThrGlyAsnPhThrLeuGlnValAlaAspLysThrGly 933  
3004 GAACCCGTAAGTCTCGACGAATTAAGCGTAGTGGAAGGAAACACACAC 3053  
934 GluPro...AsnHisAsnGluLeuThrLeuPheAspAla.....SerLy 947  
3054 ACCGCTCCGAAATCTTAATTCACCTGCAACACGACAGTCGATG 3103  
947 sAlaGlnArgAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964  
3104 CCGCGCATCGCTTATCAGCTTATCGGCAACAGCGGAGTTCGCCCTG 3153  
964 euGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeu 980  
3154 CATAATCCG..... 3162  
981 TyrAsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIl 997  
3162 ..... 3162  
997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014  
3162 ..... 3162  
1014 snGluIleAlaArgValAspGluAlaProValProProAlaPro 1030  
3162 ..... 3162  
1031 AlaThrProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGl 1047

3162 ..... 3162  
1047 uSerLysThrValGluLysAsnGlnAspAlaThrGluThrThrAlaG 1064  
3163 .....GTCAAGAACAAAGAG 3177  
1064 lnAsnArgGluValAlaLysGluAlaLysSerAsnValLysAlaAsnThr 1080  
3178 CTTTCGCACAACTCGCAAGCGGGA...GAAACAGAGCGCGCTTGAC 3224  
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097  
3225 GGCAAACAGGACCACTTCCGCCCAACACAGCGGGAAGAGAACAG 3274  
1097 rThrGluThrLysGluThrAlaThrValGluLysGluGluLysAlaLysV 1114  
3275 CGCAAGC.....CTTGACGCGCTGATTGG 3300  
1114 aGluThrGluLysThrGlnGluValProLysValThrSerGlnValSer 1130  
3301 GCCGGCGCAATGCCACCGAAAGGAGAGAGTGTGCCGAAACCGCGCG 3350  
1131 ProLysGlnGluSerGluThrValGlnProGlnAlaGluProAlaAr 1147  
3351 GCAGCGCAGCGGGGAAATGCCGCAATTATGCAGCGGAGAGAGAGAAA 3400  
1147 g.....GluAsnAspProThrValAsnIleLysGluProGlnS 1160  
3401 AACGGTGAGCGGATTAAGACACCGCTTG...GCGAAACAGCGCGAA 3447  
1160 erGlnThrAsnThrAlaAspThrGluGlnProAlaLysGluThrSer 1176  
3448 GCG..... 3450  
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193  
3451 .....GAAACCGCGCGCTACCGCGCTTCC 3478  
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209  
3479 CC.....CGCGCGCGCGCG 3492  
1209 roThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225  
3493 GCCCGCGCGGATTGCGG...CAACCGCAGCGCCCAACCGCAACCCAC 3539  
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242  
3540 GCAGCGC.....GACCTGATCAGCGCTTATGCAATA 3571  
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259  
3572 CGCGTTTGAAGTAATTTCCGCGC.....ACGCTCAACAGC 3606  
1259 laValLeuSerAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275  
3607 GTTTTCGCGCTACAGGACGAATTGGACCGCGTG...TTTGGCGAAGACCG 3653  
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1292  
3654 CGCAACCGCGTTTGGACAACGCGCATCCGGGACACCAACACACTACCGTT 3703  
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSerS 1309  
3704 CGCAAGATTTCGCGCGCTTACCGCCCAACAAACCGACTCGCGCAATCGGT 3753  
1309 erSerGlnTyrArgArgPheSerLysSerThrGlnThrGlnLeuGly 1325  
3754 ATGCAGAAAAACCTCGCGAGCGGG...CGCGTCGCGCATCCTGTTTCCGA 3800  
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyValPheThrTy 1342









422 SerAspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLys 438

1335 TACCGTTTACTGGAAAGTAAGCCGTGGCAAAACACGCCCTGTCTCAAAA 1384  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
438 sThrValThrTrpLysValHisAsnProLysSerAspArgLeuAlaLysI 455  
1385 TCGCACAAGCACGCTGCTGGTTCGAAGCCAAAGGGGAAAAACCAGGCTCG 1434  
leGlyLysGlyThrLeuIleValGluGlyLysGlyLysGlyLysGlySer 471  
1435 GTCACGCTGGCGCAGCGTAAGATCATCTTTAGATCAGCAGGCGGAGATCA 1484  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAla 488  
1485 AGGCAAAAAAACAAGCCCTTAGTGAAATCGCTTGGTCAGCGCAGGGAGA 1534  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
488 nasLysValLysAlaPheSerGlnValGlyIleValSerGlyArgSerT 505  
1535 CGGTGCAACTGAATGCCGATAATCACTTCAACCCGACAACTCTATTTC 1584  
505 hrValValLeuAsnAspAspLysGlnValAspProAsnSerIleTyPhe 521  
1585 GCCTTCGCGCGCAGCGTGTGGATTGTAACGGGCATTCGCTTTCGTCCTCA 1634  
522 GlyPheArgGlyArgLeuAspAlaAsnGlyAsnLeuThrPheGI 538  
1635 CCGCATTCAAAATACCGATCAAGGGCGGATGATTGTCAACCAACCAATCAAG 1684  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
538 uHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnThS 555  
1685 ACAAGAATCCACCGTTTACCATTACAGGCAATAAAGATATTACT..... 1728  
555 erLysThrSerThrValThrIleThrGlyGluSerLeuIleThrAspPro 571  
1728 ..... 1728  
572 AsnThrIleThrProTyrrAsnIleAspAlaProAspGluAspAsnProTy 588  
1728 ..... 1728  
588 rAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrrLeuAsnLeuGluA 605  
1728 ..... 1728  
605 snTyrrThrTyrrAlaLeuArgLysGlyAlaSerThrArgSerGluLeu 621  
1729 ..... ACAACCGCGCAATAAC..... 1743  
622 ProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrrMetGlyLysTh 638  
1744 ..... AACAACTTGGATPAGCAAAAAAG 1765  
638 rSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgM 655  
1766 AAATTGCTTACACGCTTGGTTCGCGAGAGAAGATGCCAACCAACCAAGAAC 1815  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
655 etAsnGlyPheAsnGlyTyrrPheGlyGluGluGly...LysAsnAsn 670  
1816 GGCGCGCTCAATCTGAATTACCAACCGGAAGAGCGGATCGCACTTTACT 1865  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
671 GlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPheLe 687  
1866 GCTTTCCGCGCGCAAAATTTAAACGGCAATATATCCGCAACCAACCAAGCA 1915  
687 uLeuThrGlyLysThrAsnLeuAsnGlyAspLeuLysValGluLysGlyT 704  
1916 ACTGTTTTTCAGCGGACAGCAGCCGCGCCTACATCATTTAGGA 1965  
704 hrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 720  
1966 AGCGGCTGGTCAAAATGCAAGGT.....ATCCCAACAAGGAGAAATCGT 2009  
721 IleSerSerThrLysAspGlnHisPheAlaGluAsnAsnGluValVa 737



2010 GTGGGACACGATTGGATCGACCGCACATTTAAAGCGGAAACTTCCATA 2059  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
737 lValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnV 754  
2060 TTCAGGGGGACACAGCGGTGGTTCC...CGCAATGTGGCAAGTGGAA 2106  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
754 alThrAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThr 770  
2107 GGGGATTGGCATTAAAGCAATCAGCGCAAGCAGATTTCGGTGTGCGACC 2156  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
771 SerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyr 785  
2157 GCATCAAGACACACATCTGTACAGCTCGGACTGGAGCGGTCTGACAA 2206  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
786 .LysAlaGlyAspThrValCysValArgSerAspTyrThrGlyTyrValt 802  
2207 GTTGTTACCGAAAAACCATTCACGACGATAAAGTATGTTCTTCATTGAGC 2256  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
802 hrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsn 817  
2257 AAGACCGACATCAGAGCGCAATGTACAGCTTTCGGGATCACGCTCATTTAAA 2306  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
818 AlaThrAsnValSerGlyAsnValAsnLeuSer 828  
2307 TCTCAGAGGACTGCCACACTCAACGGCAATCTTAGTGCAGCGGAGACA 2356  
828 ..... 828  
2357 CGCACTATACGGTTACGGCAACCGCCACCCAAACCGGCAACCTCACCCCTC 2406  
828 ..... 828  
2407 GTGGGCAATGCCAAGCAACATTAATCAAGCCACATTAAGCGGCAACAC 2456  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
829 ..GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrI 844  
2457 ATCGGCTTCGGACAAATGCTTCATTTAATCTAAGCAACACCGCGTACAAA 2506  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
844 eSerGlyThrGlyAsnSerGlnValArgLeuThrGluAsn 857  
2507 ACGGCAGTCTGAGCGTTCCGACACAGCTAAGGCAACGCTAAGCCATTC 2556  
857 ..... 857  
2557 GCACCAACGGCAATGTCTCCTAGCGGATAAGCGAGTATTCATTTGA 2606  
857 ..... 857  
2607 AACACCCGCTTACCGGAAATATCAGCGCGGCAAGGATACGGCATTAC 2656  
857 ..... 857  
2657 ACTTAAAGACAGGAATGACGCTCGCTCGGCGCACGGAATTAGGCAAT 2706  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
858 .....SerHisIlePheHisLeuThrGlyAspSerAsnValAsnGln 870  
2707 TTAACCTTTGACACGCCACCATTAACATCAATTCGCGCTATCGACACGA 2756  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
871 LeuAsnLeuAspLysGlyHisIleHisLeuAsnAla.....GlnAsnAs 885  
2757 TCGGCGAGCGGCAACCGCGGAGTGGCGGACATCGCGCGCGCGCGCGT 2806  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
885 pAlaAsnLysValThr 891  
2807 CGCGCGTTCCTTATATCCGTTACCGCGCCCAACTTCGCGCAGATCCCGT 2856  
891 ..... 891  
2857 TTCAACACGCTGACGCTAAACGGCAAAATTGACGCTGAGGAAACATCCG 2906  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPhe 907

2907 CTTTATGTCGGAACCTCTTCGGCTACCGCAGCGGCAAAATTGAAGCTGGCG 2956  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
907 rTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysValValThrL 924  
2957 AAAGTTCGGAAGGCACCTTACACCTTGGCTGTCAACAATACCGCAACGAA 3006  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
924 ysSerAlaThrGlyAsnPheThrLeuGlnValAlaAspLysThrGlyGlu 940  
3007 CCGGTAAGTCTCGACCAATTCAGCGTAGTGGNAGGAAAGACACACACC 3056  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
941 ProThrLys...AsnGluLeuThrLeuPheAspAlaSerAsnAlaThr 955  
3057 GCTGTCCGAAATCTTAATTTACCTCTCAAAACGACACACGTCGATGCG 3106  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
956 ....ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAspLeuG 971  
3107 GCGCATGCGGTTATCAGCTTATCCGCAAGACGCGAGTTCGCGCTGCAT 3156  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
971 lYalaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeuTyr 987  
3157 NATCCG..... 3162  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004  
3162 ..... 3162  
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021  
3162 ..... 3162  
1021 luGluIleAlaArgValGluThrProValProProAlaProAlaThr 1037  
3162 ..... 3162  
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGluSerLy 1054  
3163 .....GTCAAAGAACAGAGCTTCGCAAACTCCGCAAACTCCGCAAGCGG 3202  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsnG 1071  
3203 GAGAA..... 3207  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1071 lyGluValAlaGluLysProSerValLysAlaAsnThrGlnThr 1087  
3208 .....ACAGA 3212  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrThrGl 1104  
3213 GCGCGCTTTCACGCA..... 3228  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1104 uIleLysGluThrAlaLysValGluLysGluLysAlaLysValGluL 1121  
3229 .....AAA 3231  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1121 ysAspGluIleGlnGluAlaProGlnMetAlaSerGluThrSerProLys 1137  
3232 CAGGCACAACTTCGCGCAAA..... 3252  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1138 GlnAlaLysProAlaProLysGluValSerThrAspThrLysValGluG 1154  
3252 ..... 3252  
1154 uThrGlnValGlnAlaGlnProGlnThrGlnSerThrThrValAlaAla 1171  
3253 .....CAACAGGCG 3261  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1171 laGluAlaThrSerProAsnSerLysProAlaGluGluThrGlnProSer 1187  
3262 GAAAAACACACCGCAAGCCCTGACGCTGATTCGCGCGCGGCGCA 3311  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :  
1188 GlulysThrAsnAlaGluProValThrProValValSerLysAsnGlnTh 1204  
3312 T.....G 3313





1661 CGATGATTGTCAACCAATCAAGACAAAGAAATCCACCGTTACCAATTACA 1710  
|| ..... ||| ..... :||| ..... |||  
545 laArgLeuValAlaHisSerThrSerLysHisSerThrValThrIleThr 561  
|| ..... ||| ..... :||| ..... |||  
1711 GGCAATAAAGATATTACTACACCGGCAAT ..... 1740  
|| ..... ||| ..... :||| ..... |||  
562 GlyAspAsnLeuIleThrAspProAsnAsnValSerIleTyrTyrVally 578  
1740 ..... 1740  
578 sProLeuGluAspAspAsnProTyrAlaIleArgGlnIleLysTyrGlyT 595  
1740 ..... 1740  
595 yrGlnLeuTyrPheAsnGluGluAsnArgThrTyrTyrAlaLeuLysLys 611  
1740 ..... 1740  
612 AspAlaSerIleArgSerGluPheProGlnAsnArgGlyGluSerAsnAs 628  
1741 ..... A 1741  
628 nSerTyrLeuTyrMetGlyThrGluLysAlaAspAlaGlnLysAsnAlaM 645  
1742 ACAACAACCTTGGATAGCAAAAGAAATTCGCTACACGGTTGGTTTCGC 1791  
|| ..... ||| ..... :||| ..... |||  
645 etAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly 661  
1792 GAGAAGATGCAACCAACCAACGACGGGGTCAATCTGAATTACCAACC 1841  
|| ..... ||| ..... :||| ..... |||  
662 GluGluGluGly...LysAsnAsnGlyAsnLeuAsnValThrPheLysG 677  
1842 GGAAGAAGCGGATCGCACTTTACTGCTTTCGGGGGGAACAAATTTAAACG 1891  
|| ..... ||| ..... :||| ..... |||  
677 LysSerGluGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnG 694  
1892 GCAATATCACGCAAAACCAACGCAACTGTTTTTCAGCGGACGACGAC 1941  
|| ..... ||| ..... :||| ..... |||  
694 LysPheLeuAsnValGlnGlnGlyThrLeuPheLeuSerGlyArgProThr 710  
1942 CGCACGCGCTACAATCAITTAGGACGGGTGGTCAAAATGGAAGT... 1989  
|| ..... ||| ..... :||| ..... |||  
711 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspSerHi 727  
1990 ....ATCCCAACGAAGGAGAAATCGTGGGACAAACGATTGGATCGACCGCA 2035  
|| ..... ||| ..... :||| ..... |||  
727 sPheSerGluAsnAsnGluValValValGluAspAspTrpIleAsnArg 744  
2036 CATTAAAGCGGAAACCTTCCATATTCAGGCGGACAGCGGTGTTC 2085  
|| ..... ||| ..... :||| ..... |||  
744 snPheLysAlaThrAsnIleAsnValThrAsnAsnAlaThrLeuTyrSer 760  
2086 ....CGCAATGTCGCAAGTGAAGCGCATTTGGCATTTAAGCAATCAACG 2132  
|| ..... ||| ..... :||| ..... |||  
761 GlyArgAsnValGluSerIleThrSerAsnIleThrAlaSerAsnAla 777  
2133 CCAAGCAGTTTCGGTGTGCGACCGCATCAAGCCACACCAATCTGTACAC 2182  
|| ..... ||| ..... :||| ..... |||  
777 LysValHisIleGlyTyr.....LysAlaGlyAspThrValCysVala 792  
2183 GTTCGGACTGGACGGTCTGACAACTGTGACGAAACCAATACCGAC 2232  
|| ..... ||| ..... :||| ..... |||  
792 rgSerAspTyrThrGlyTyrValThrCysThrThrAspLysLeuSer... 807  
2233 GATAAGTGATTGCTTCATTGAGCAAGACCGACATCAGAGGCAATGTCCAG 2282  
|| ..... ||| ..... :||| ..... |||  
808 AspLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAsnValas 824  
2283 CTTTCGGCATCGCTCAATTAATCTCACAGGACTTGCACACTCAACG 2332  
|| ..... ||| ..... :||| ..... |||  
824 nLeuThrGluSerAlaAsnPheValLeu...GlyLysAlaAsnLeuPheG 840  
2333 GCNAATTTAGTCAGCGGAGACACGCATATACGGTTACGGCAACGCC 2382

840 lyThrIleGlnSerArgGlyAsnSerGlnValArgLeuThrGluAsn... 855  
2383 ACCCAAAACGGCAACCTCAGCCTCGTGGCAATGCCAAGCAACATTTAA 2432  
855 ..... 855  
2433 TCAAGCCACATTAAACGGCAACACATCGGCTTCGGACAATGCTTCATT 2482  
855 ..... 855  
2483 ATCTAAGCAACAAACGCCGTACAAAACGGCAGTCTGACGCTTTCGCACAC 2532  
855 ..... 855  
2533 GCTAAGGCAACGTAAGCCATTCCGCACTCAACGGCAATGTCTCCTAGC 2582  
855 ..... 855  
2583 CGATAAGGCAGTATTCATTGTTGAAAACAGCCGCTTTACCGGAAAAATCA 2632  
855 ..... 855  
2633 CGCGCGCAAGGATACGGCATTTACACTTAAAAGACACGAATGGACGCTG 2682  
856 ..... SerHisTrpHisLeu 860  
2683 CGCTCGGCACGGAATTAGGCAATTTAAACCTTGACACGCCACCATTTAC 2732  
|| ..... ||| ..... :||| ..... |||  
861 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleHi 877  
2733 ACTCAATTCGCGCTATCGACACGATCGCGGCGGCAACCGGCGAGTG 2782  
|| ..... ||| ..... :||| ..... |||  
877 sLeuAsnSerAlaAspAsnSerAsn ..... 885  
2783 CGCAGATGCGCGCGCGCGCTTCGCGGTTCCCTATTATTCGCTACG 2832  
886 ..... AsnValThr 888  
2833 CGCCCAACTTCGCGCAGAAATCCCGTTTCAACACGCTGACGTAACGCAA 2882  
|| ..... ||| ..... :||| ..... |||  
889 .....LysTyrAsnThrLeuThrValAsn...Se 897  
2883 ATTGAACGTCAGGCAACATTCGCTTTATGCGGAACCTTCGCGCTACC 2932  
|| ..... ||| ..... :||| ..... |||  
897 rLeuSerGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysG 914  
2933 GCAGCGCAAAATTTAAAGTCGGGCAAGTCCGAAGGCACCTTACACCTTG 2982  
|| ..... ||| ..... :||| ..... |||  
914 lnglyAspLysValValThrLysSerAlaThrGlyAsnPheThrLeu 930  
2983 GCTGTCAACAATACCGCAACACCCGTAAGTCTCGAGCAATTCAGGT 3032  
|| ..... ||| ..... :||| ..... |||  
931 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrLe 946  
3033 AGTGGAGGAAACACACACACCGCTGTCGGAATCTTAATTTCCACC 3082  
|| ..... ||| ..... :||| ..... |||  
946 uPheAspAla.....SerLysAlaGlnArgAspHisLeuAsnValSerL 961  
3083 TGCAAAACGAACACGTCGATGCGCGCGCATGGGTTATCAGCTTATCCGC 3132  
|| ..... ||| ..... :||| ..... |||  
961 euValGlyAsnThrValAspLeuGlyAlaThrLysTyrLysLeuArgAsn 977  
3133 AAAGACGGCGAGTTCGCCCTGCATAATCCG ..... 3162  
|| ..... ||| ..... :||| ..... |||  
978 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnG 994  
3162 ..... 3162  
994 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1011  
3162 ..... 3162

```

1011 alProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1027
3162 ..... 3162
1028 ProValProProAlaProAlaThrProSerGluThrThrGluThrVa 1044
3162 ..... 3162
1044 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnA 1061
3162 ..... 3162
1061 sAlaThrGluThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1077
3163 .....GTCAGAACAGAGCTTCCGACAACTCGGCAAGCGGA... 3204
1078 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySe 1094
3205 .GAACAGAGCCCTTACGGCAAAACAGGACACACTTCCGCGCAAAAC 3253
1094 rGluThrLysGluThrGlnThrThrGluThrLysGluThrAlaThrValG 1111
3254 AACAGCGGAAAGACCAACGCCAAGC..... 3282
1111 LuLysGluGluLysAlaLysValGluThrGluLysThrGlnGluValPro 1127
3283 ...CTTGACGCGTGTATTGCGCGCGGCGCAATGCCACCGAAAGCAGA 3329
1128 LysValThrSerGlnValSerProLysGlnGluGlnSerGluThrValGI 1144
3330 AAGTGTTCGCGAACCGCGCGGAGGAGCGGGGAAATTCGCGGCAATTA 3379
1144 nProGlnAlaGluProAlaArg.....GluAsnAspProThrV 1157
3380 TGACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3429
1157 alAsnIleLysGluProGlnSerGlnThrAsnThrThrAlaAspThrGlu 1173
3430 TTG...GCGAACAGCGCGCAAGCG..... 3450
1174 GlnProAlaLysGluThrSerSerAsnValGluGlnProValThrGluse 1190
3451 .....GAAACCC 3457
1190 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThrT 1207
3458 GGCGCGGTACACCGCGCTTCCCG..... 3480
1207 hrProAlaThrThrGln...ProThrValAsnSerGluSerSerAsnLys 1222
3481 .....CGCGCGCGCGCGCGCGCGCGGATTTCCG...CAACCGCA 3518
1223 ProLysAsnArgHisArgArgSerValArgSerValProHisAsnValGI 1239
3519 GCCCAACCGCAACCCACCGCAGCGC.....GACC 3550
1239 uProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuLysAspL 1256
3551 TGATCAGCGCTTATGCCAATAGCGGTTTTCAGTGAATTTTCCGCGC... 3594
1256 euThrSerThrAsnThrAsnAlaValLeuSerSerAspAlaArgAlaLysAla 1272
3595 .....ACGTCACAGCGGTTTTCGCGCGCGCGCGCGCGGATTTGACCG 3635
1273 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGI 1289
3636 CGTG...TTGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 3682
1289 nLeuGluMeLAsnAsnGluGlyGlnThrAsnValThrPValSerAsnThrS 1306
3683 GGGACACCAAACTACCTGTCGCAAGATTTCCGCGCGCTACCGCGCAACAA 3732
1306 erMetAsnLysAsnTyrSerSerSerGlnTyrArgArgPheSerSerLys 1322

```

```

3733 ACCGACCTCGCCCAATCGGTATGCAGAAAAACCTCGCGCAGCGGG...CG 3779
1323 SerThrGlnThrGlnLeuGlyTrpAspGlnThrIleSerAsnValGI 1339
3780 CGTCGCGCATCTGTTTCGCACACCGGACCGGAAACACCTTCGACGAGC 3829
1339 nLeuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspLysA 1356
3830 GCATCGGCAACTCGCGACGCGTTCGCCACCGGTGCGCTTTTCGGGCAATAC 3879
1356 laThrSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysTyrTyr 1372
3880 GCATCGGCGAGTTCGACATCGGCATCAGCGCGGCGCGGCTTTAGTAG 3929
1373 AlaAspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGI 1389
3930 CGGACGCTTTTCAGACGCGCATCAGAGGCAAAATCCGCGCGCGCTGCTGC 3979
1389 nSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406
3980 ATTACGGCATTTCAGCAAGATACCGCGCAGGTTCGCGGATTCGGCATC 4029
1406 nPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIle 1422
4030 GAACCGCATCAGCGCAACCGCTATTTCGTCCCAAAAGCGGATTACCG 4079
1423 ThrProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439
4080 ATACGAAACGTCATATATCGCCACCGCGGCTTCGCTTCAACCGCTACC 4129
1439 aLeuAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaP 1456
4130 GCGCGGCGATTTCAGCATTCATTCAACCGCGCGCAACACATTTC 4179
1456 heAlaGlnValAspLeuSerTyrThrTyrHisLeuGlyGlu...PheSer 1471
4180 ATCAGCGCTTATTTCAGCGTGTCTATACCGATCGCGTTCGCGCAAGT 4229
1472 ValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlySerGI 1487
4230 CGAACCGCGCTCAATATACCGCGGCTTCGCGCAGGATTCGCGCAAAACCC 4279
1487 yLysIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnGlnG 1504
4280 CGATCGCGGATTCGCGGCTAACCGCGGAAATCAAGGTTTCACGCTGCC 4329
1504 nGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1520
4330 CTCCACGCTGCGCGCAAGGCGCGCAATTCGCGCAGGATTCGCGCAAAACCC 4379
1521 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAl 1537
4380 GGGCATCAATATAGGCTAC 4398
1537 aGluLeuLysLeuSerPhe 1543

```

seq\_name: pir2:A81018

seq\_documentation\_block:

serine-type peptidase NMB1998 [imported] - Neisseria meningitidis (strain MC58 serogroup B)  
 C:Species: Neisseria meningitidis  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C:Accession: A81018  
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R. A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755  
 A:Accession: A81018  
 A>Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1431 &lt;TE&gt;

A;Cross-references: GB:AE002549; GB:AE002098; NID:g7227258; PIDN:AAE42325.1; PID:g722726

A;Experimental source: serogroup B, strain MC58

C;Genetics:

A;Gene: NMB1998

## alignment\_scores:

Quality: 1585.50	Length: 1668
Ratio: 1.829	Gaps: 50
Percent Similarity: 51.978	Percent Identity: 28.118

## alignment\_block:

US-09-303-518D-653 x A81018 ..

Align seg 1/1 to: A81018 from: 1 to: 1431

```
55 GGCGCATCCGCTTCTCGCCGCTTACTAGCCATATGCTGCTGTCGG 104
|||||
17 GlyLeuPheAlaValSerProAlaTyrSerSerileValArgasn..... 31
105 CATTCTGCCCAAGCCGCGGGGACACACTATTTCGGCAATCACTACC 154
|||||
32 .....AspValAspTyrG 36
155 RATACTATCGGACTTTCGCCAAATAAGCAAGTTTGCAGTCGGGCG 204
|||||
36 InTyrPheArgAspPheAlaGluasnLysGlyAlaPheThrValGlyAla 52
205 AAAGATATTAGGTTTACACAAAAAGGGAGTTGTCGGCAATGAT 254
|||||
53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69
255 GACGAAACCCGATGATGATTTTCTGCTGTCGCT...AACGGCG 301
|||||
69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86
302 TGGCGCATTCGGCGGCATCAATATATGTCAGGTGGCAATCAAGCGC 351
|||||
86 LeAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102
352 GCTATAACATGTTGATTTGGTGGCGGAGGAAACCAATCCGATCAGCA 401
103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGluGlu 119
402 CCGCTTTCTTACCAATTTGTAAGAAATAATTAAGACGGGACTA 451
119 nAlaTyrThrTyrArgLeuValSerArg.....A 129
452 ACGGCCATCTTATGCGCGCATATATATGCGCGTTTTCACAAATTT 501
129 snProHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeu 144
502 GTCACATGACGAACCTGTTGAGATGACCACT..... 534
145 ValThrGluIleSerProThrAlaLeuSerSerValProLeuLeuGlyAs 161
535 ....TATATGGATGGTGAATACGCTGATTTAAATAATACCTGATC 580
161 nGlyGlnProLysAlaAsnAlaTyrLeuAspThrAspArgPheProTyrP 178
581 GGTTCGATCGAGCAGCAGACAAATATGCGGCTCTGATCAGACAGAA 630
178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193
631 CCAATAACCGGAAAGTTCATATATATGCAAGCGCATATCTTGCT 680
194 .....ThrArgThrArgThrAlaProAlaTyrGlnTyrLe 205
681 CGTCGCTGCAATACCTTTGCAAAATGGATGAGTGGTGGCAGACATCA 730
205 uThrGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeuL 222
```

```
731 ACTTAGGTAGCGAAATAAATTAACATATGCTGTTTTCACCAACA 780
|||||
222 euValGlyGly...SerLeuThrAspGlnPro.....LeuAsnThr 234
781 GGAGGCTCATTTGGCGACAGTGGCTACCAATGTTTATCTATGATGCCCA 830
|||||
235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysH 251
831 AAAGCAAAAGTGTAAATTAATGGGTATTCGAAACAGCAAGCAACCCCTATA 880
|||||
251 sGluAsnArgTrpValLeuAlaGlyValLeuSerThr.....TyrA 265
881 TAGGAAAAAGCAATGCTTCCAG.....CTAGTTCTGTAAGATTGGTTC 924
|||||
265 laGlyPheAspAsnPhePheAsnLysTyrIleValThrGlnProGluPhe 281
925 TATGATGAAATCTTCTCGAGATACCCATTCAGTATCTTACGAAACCACA 974
|||||
282 IleArgSerThrIleArgGln.....TyrGluThrAr 292
975 T.....CAAAATGGGAATACTTTTAAACGACAAATAA 1009
292 gLeuAspValGlyLeuThrThrAsnGlnLeuIleTrpArgAspAsnGlyA 309
1010 ATGGCGCAGGAAAAATCGATGCGCAAAACACTATTTCTTACCTTAT 1059
|||||
309 snGlyAsnSerThrLeuGlnGlyLeuAsnGluArgIleThrLeuPro... 324
1060 AGATTAAAAACAGCAACCGTTCATTTGTTTATGTTTATCCGAGAC 1109
325 .....IleAlaAsnProSerLeuAla..... 331
1110 AGCAAGAGAACCTGTTTATCATCTGCGAGTGGGTCAACAGTTATCGAC 1159
332 .....P 332
1160 CCAGACTGAATAATGGAGAAATATTTCTTTTATTGACAAAGGAAAA... 1206
|||||
332 roGln...AsnAspSerArgHisMetProSerGluAspAlaGlyLysThr 347
1207 .....GGTGAATTGATCTTACCAGCAA 1229
348 LeuIleLeuSerSerArgPheAspAsnLysThrLeuMetLeuAlaAsp 364
1230 CATCAACCAAGCGCGGGTGTGTTGTTGAGGTAAATTTTACGCTCT 1279
|||||
364 nIleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAsnPheThrVal 381
1280 CGCTAAAAACAAACGTCGAGCGGGCGGTTCATATCATCATGAT 1329
|||||
381 aGlyLysAsnHis...ThrTrpGlnGlyAlaGlyValIleValAlaAsp 396
1330 GGAGTACCTGTTACTTGGAAAGTAAACGGGTGGCAACAGCCGCTGTC 1379
|||||
397 GlyLysArgValPheTrpGlnValSerAsnProLysGlyAspArgLeu 413
1380 CAAATCGCAAGGACGCGCTGTTCAAGCAAGGGGAAACCAAG 1429
|||||
413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGln 430
1430 GCTCGTACGCGTGGCGAGGTAAAGTCACTTATGATCAGCAGCGGAC 1479
|||||
430 LyAspIleSerIleGlyGlyValValValLeuAlaGlnLysAlaAla 446
1480 GATCAGGCAAAACAGCCTTTAGTAAATCGCTTGGTGGCAGCGGAC 1529
|||||
447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGly 463
1530 GGGAGCGTGCCTGAATGCCGATATCACTCAACCCCGACAAACTCT 1579
|||||
463 gGlyThrAlaValLeuAlaAspSerGlnGlnIleLysProGluAsnLeu 480
```







A:Cross-references: GB:M87489; NID:q148906; PIDN:AA24966.1; PID:q148907

A:Experimental source: strain HK715

A:Note: Sequence extracted from NCBI backbone (NCBIP:97282)

C:Superfamily: IgA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

#### alignment\_scores:

Quality: 1583.50 Length: 1867  
Ratio: 1.679 Gaps: 48  
Percent similarity: 50.509 Percent identity: 25.174

#### alignment\_block:

US-09-303-518D-653 x A41859 ..

Align seg 1/1 to: A41859 from: 1 to: 1702

```
64 CGTCTCTCGCGCGTACTTAGCATATGCTGCTCGTTCGGCATTCCTGCC 113
:|||||: :|||||: :|||||: :|||||: :|||||: :|||||:
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaLeuThrValAlaLeuThrPr 21
114 CCAAGCCCGCGGGGACACACTATTTCGGCATCAACTACCACTACTATC 163
| :|||||: :|||: :|||||: :|||||: :|||||: :|||||:
21 oTyrThrGluAlaLeuValArgAspValAspTyrGlnIlePheA 38
164 GCGACTTTGCCGAAATAAGGCAAGTTTCAGTCGCGGGGAAAGATATT 213
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
38 rGaspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
214 GAGTCTTACAAAAAAGGGAGTGTTCGGCAATCGATGACGAAAGC 263
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG1 71
264 C...CGATGATTGATTTTCTGTGTATCG...CGTAACGGCGTGGCGG 307
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
71 YileProMetIleAspPheSerValValAspValAspLysArgIleAlat 88
308 CATTGGCGGCGATCAATATATCTGAGCGTGCACAT...AACGGCGGC 354
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATAACAATGTTGATTTTGTG.....GCGAGGGAAGCAA 389
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnGlyAs 121
390 TCCCGATCAGACCGC.....TTTCTTACCAATTCG 421
| :|||||: |||||: |||||: |||||: |||||: |||||: |||||:
121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138
422 TGAAGAATAATATTATAAGCAGGACTAACGGC..... 456
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
138 alGluLysAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
457 .....CATCCTTATGCGCGGATTCATATCCCGCTTGA 494
155 GluAspGlnThrGlnLysArgGluAspTyrMetProArgLeuAs 171
495 CAAATTGTACAGATGCAGACCTGTTGAG...ATGACAGTTATATGG 541
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188
542 ATGGGTGGAATACGCTGATTAATAATACCTGATCGTGTTCGAATC 591
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
592 GGAGCAGGAGCAATATTGGCGGCTGTGATGAGACCAA..... 630
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
205 GlySerGlnThrPheIleTyrLysLysGlyAspAsnTyrSerLeuI1 221
631 .CCCAATAACCGGAAAGTTTCATATCAT.....ATTGCAAGCG 667
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
221 eLeuAsnAsnHisGluValGlyClyAsnAsnLeuLysLeuValGlyAspA 238
```

```
668 CATATCTTTGGCTCGTGGTGGCAATACTTTGCACAAAAATGATCAGGT 717
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
238 lAtyThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGluAsn 254
718 GTGGCAGACAGTCACTTAGTAGCCGAAAAAATAACATAGC...CCATA 764
: |||||: :|||||: :|||||: :|||||: :|||||: :|||||:
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProly 271
765 TGGTTTATTTTACCAACAGGA.....GGCTCATTTTGGCGACA 799
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyAspS 288
800 GTGGCTCACCATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTAATT 849
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
288 erGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeuPhe 304
850 AATGGGGTATTGCAACAGGCAACCCCTATATAGGAAAAAAGCAATGGCTT 899
|||: |||: |||: |||: |||: |||: |||: |||:
305 LeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSerTrpGlnG1 321
900 CCAGCTAGTTTCGTAAGATTTGTTCTATGATGAAATCTTTGCTGGAGATA 949
: |||||: :|||||: :|||||: :|||||: :|||||: :|||||:
321 uTrpAsnIleTyrLysProGluPheAlaLysThrValLeuAspLysAspT 338
950 CCCATTTCAGTATTCACGAACACATCAAAATGGGAAATACTTTTTTAAC 999
|||: |||: |||: |||: |||: |||: |||: |||:
338 hr..... 338
1000 GACATATAATGCGCGCAGGAAAAATCGATGCCAACATAACTACTATTC 1049
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
339 .....AlaGlySerLeuThrGlySerAsnThrGlnTyrAs 350
1050 TCTACCTTATAGATTAAAAACAGCAACGCTT.....CAATTGT 1087
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
350 nTrpAsnProThrGlyLysThrSerValIleSerAsnGlySerGluSerL 367
1088 TTAATGTTTCTTTATCCGAGACAGCAAGAACCTGTTTATCATGCTGCA 1137
: |||||: :|||||: :|||||: :|||||: :|||||: :|||||:
367 euAsnValAspLeuPheAspSerSerGlnAsp..... 377
1138 GGTGGGCTCAACAGTTATCGACCCAGACTGAATAATGAGAAAAATATTC 1187
: |||: :|||: :|||: :|||: :|||: :|||: :|||:
378 .....ThrAspSerLysLysAsnAsnHisGlyLysSerValTh 390
1188 CTTTATTGACAAAGAAAGGTGAATTCATCTTACCAGCAACATCAACC 1237
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
390 rLeu.....ArgGlySerGlyThrLeuThrLeuAsnAsnAlaAspG 405
1238 AAGCGCGCGGCGTTGTTTATTTTTCAGGGTAATTTTACGGTC...TCGGCT 1284
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
405 InGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThr 421
1285 AAAACAACAGAACGTGCGCAAGCGCGCGCTTCATATCATGATGGCAG 1334
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
422 SerAspSerThrTrpLysValHisAsnProLysSerAspArgLeuAlaLys 438
1335 TACCGTTTACTTGGAAAGTAAACGCGTGGCAACGCGCTGTCACAAA 1384
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
438 sThrValThrTrpLysValHisAsnProLysSerAspArgLeuAlaLys 455
1385 TCGGCAAGGCGCGTGTGTTTCAAGCAAAAGGGGAAAAACCAAGGCTCG 1434
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
455 leGlyLysGlyThrLeuIleValGluGlyLysGlyGluAsnLysGlySer 471
1435 GTCAGCGTGGCGGCGGTAAAGTCAATCTTAGATCAGACGGCGGACGATCA 1484
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAs 488
1485 AGGCAAAAAACAACCCCTTACTGAAATCGGCTTGGTACGCGCGGAGGGA 1534
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
488 nasLysValLysAlaPheSerGlnValGlyIleValSerGlyArgSerT 505
1535 CGGTGCAACTGAATGCCGATTAATCAGTTCAACCCCGACAAACTCTATTC 1584
```











987 erProAlaAlaAsnThrAlaSerGlnAlaGlnLysAlaThrGlnThrAsp 1000  
3214 GCCGCTTTCAGCGCAAAA.....  
1004 GlyAlaGlnIleAlaLysProGlnAsnIleValValAlaProProSerPr 1020  
3232 .....CAGGCACAACCTTCCGCCCAACACACAGCGGAGAGACA 3271  
1020 ogInAlaAsnGlnAlaGluAlaLeuArgGlnGlnAlaLysAlaGluG 1037  
3272 ACGGC.....CAAGCCTTGACCGCGCTGATTCGGCGCGCGCAAT 3312  
1037 InValLysArgGlnGlnAlaAlaGluAlaGluLysValAlaArgGlnLys 1053  
3313 GCCACCGAAAGGCGAAGGTGTTCGCGAAGCGCGCGCGCGAG..... 3354  
1054 AspGluGluAlaLysArgLysAlaAlaGluIleAlaArgGlnGluGlu 1070  
3355 .....CGAGCGGGGAAATGCGGCATTTATCGAGCGGAGAGAGA 3399  
1070 uAlaArgLysAlaAlaGluAlaLysGlnLysAlaGluAlaGlu 1087  
3398 AAAACGGGTGCAGCGGATAAACACACCGCTTGGCGAACAACGCCGAA 3447  
1087 rGlyAlaAlaArgL.....LeuAlaArgGlnLysAla 1097  
3448 CGGAAACCGCGCGCTACACCGCTTCCCGCGCGCGCGCGCGCGCG 3497  
1098 GluGluAlaSerHisGlnAlaAsnAlaLysProLysArgArgArgAr 1114  
3498 CGGGATTTCGCGCAACCG..... 3516  
1114 gAlaIleLeuProArgProAlaProValPheSerLeuAspAspTyra 1131  
3516 ..... 3516  
1131 sPalLysAspAsnSerGluSerSerIleGlyAsnLeuAlaArgValIle 1147  
3516 ..... 3516  
1148 ProArgMetGlyArgGluLeuIleAsnAspTyrGluGluIleProLeuG 1164  
3516 ..... 3516  
1164 uGluLeuGluAspGluAlaGluGluAlaArgGlnAlaThrGlnPheH 1181  
3516 ..... 3516  
1181 isSerLysSerArgAsnArgAlaIleSerSerGluProSerSerAsp 1197  
3516 ..... 3516  
1198 GluAspAlaSerGluSerValSerThrSerAspLysHisProGlnAspAs 1214  
3516 ..... 3516  
1214 nThrGluLeuHisGluLysValGluThrAlaGlyLeuGlnProArgAla 1231  
3517 ..CAGCCCCAACCGCAACCCCAACCGCGCGCGCGCGCGCGCGTAT 3564  
1231 laGlnProArgThrGlnAlaAlaAlaGlnAlaAspAlaValSerThrAsn 1247  
3565 GCAATACGGTTTCAGTGAATTTTCGCGCGCGCGCGCGCGCGT..... 3609  
1248 ThrAsnSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLe 1264  
3610 .....TTCGCGGTACAGAGCAATTCGCGCGCGCGCGTTCGCG 3646  
1264 uAspThrGlyAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAla 1281  
3647 AGACCGCGCGCAACCGCTTTGGACAACCGCGCGCGCGCGCGCGCG 3696  
1281 sPalGluLysAsnSerValTrpMetSerAsnThrGlyTyrGlyArgAsp 1297











A:Molecule type: DNA  
 A:Residues: 53-548 <LOW>  
 A:Cross-references: EMBL:X82471; NID:g732858; PIDN:CAA57854.1; PID:g732859  
 A:Experimental source: strain HF159  
 A:Accession: S61321  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 53-548 <LOW>  
 A:Cross-references: EMBL:X82472; NID:g732852; PIDN:CAA57855.1; PID:g732853  
 A:Experimental source: strain SM1027  
 C:Genetics:  
 A:Gene: Iga; NMA0905  
 C:Superfamily: Iga-specific metalloendopeptidase  
 C:Keywords: hydrolase; metalloproteinase

alignment\_scores:  
 Quality: 1515.00 Length: 1958  
 Ratio: 1.647 Gaps: 53  
 Percent Similarity: 46.987 Percent Identity: 24.974

alignment\_block:  
 US-09-303-518D-653 x A81937 ..  
 Align seg 1/1 to: A81937 from: 1 to: 1773

```

49 AAACCGCGCGCATCCGCTTCTCGCGCGCTTACTTAGCCATATGCTGTC 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LysThrLysArgPheLysIleAsnAlaIleSerLysIlePheLeuAl 18
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 GTTCGGCATTCGCGCAACCGCGCGCGGACACACTTATTCGGCATCA 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspValA 35
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 ACTACCAATACTACGCGACTTTCGCGAATAAAGCGAGTTGTCAGTC 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 spYrGlnIlePheArgAspAlaGluAsnLysGlyLysPheVal 51
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 GGGCGGAAGATATGAGGTTTACACAAAGGGGAGTTGTCGGCAA 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyAsnIleGlyAs 68
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
249 ATCGATGACAAAGCCCGATGATGATTTCTGTCGATCGCTAAC. 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 ..GGCGTGGCGGATTCGGCGGCGATCAATATATGTCGCGTGGCACA 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 rArgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 AACGGCGGC.....TATACAATGTTGATTTGGTGGCGGA 380
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ValLysGlyAspGluIleSerTyrGlyHisHisAsnGlyHisLeuAs 118
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 GGAAGCAATCCGATCAGCAGCGCTTTCTTACCAAATTTGTGAAAGAA 430
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 pValSerAsnAspGluAsn.....GluTyrArgSerValAlaGlnA 132
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 ATATATTAAGCAGGACTTAACGGCCATCCCTATGCGGC..... 471
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 snAspTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArg 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
472 .....GATTATCATATGCGCGTTCACAAATTTGTACAGATGCAGA 515
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 LeuGluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAl 164
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 ACCTGTTGAGATGACGAGTATATGATGGGTGAAA...TACGCTGATT 562
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 aProIleAlaProThrSerAlaGlyGlyValGluThrTyrLysAspL 181
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
563 TAAATAAATACCTGATCGTGTTCGATCGGACGAGGACAGCAATATGG 612
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 ysAsnArgPheSerGluPheValArgValAlaGlyThrGlnPhe... 196
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

613 CGGTCTGATGAAGACGACCAACCAATACCGCAAACTTCATATCATATTGC 662
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSe 208
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
663 AAGCCGATATCTTGGCTGCGTGGCGCAATACCTTT..... 699
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 rArgAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnV 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
700 ..GCACAAATGATCAGGTGGTGGCAGACAGTCACTTAGTAGCGCAAAA 747
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 alThrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSer 241
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
748 ATTAACATAGCCCA.....TATGGTGTGTTT 773
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 LysHisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLe 258
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
774 ACCAAGAGAGCTCATTTGGCGACAGTGGCTCACCATGTTTATCTATG 823
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 uThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrA 275
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
824 ATGCCCAAAAGCAAAAGTCTTAATATGGGTATTCGAAACAGGCAAC 873
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
275 spLysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAla 291
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
874 CCTATATAGGAAAGCAATGGCTTCCAGCTAGTTCGTAAGATTGG.. 921
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 GlyTyr.....GlnLysAsnSerTrpGln.....GluTrpAs 302
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
922 .TTCTATGATGAATCTTTTGGTGGATACCCATTCAGTATTTCTACGAAC 970
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 nIleTyrLysLysGluPheAlaAspGluLysGln..... 314
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
971 CACATCAAAATGGGAATACCTTTTAAACGACATAATAATGGCGCAGGA 1020
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 .....ArgAspAsn.....AlaGly 319
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1021 AAAATCGATGCCAAATACACACTTCTCTACCTTATAGATTAATAAAC 1070
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
320 ThrIleLysGlyTyrGlyGluHisHis.....TrpLysTh 331
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1071 AGAACCCGTTCAATTTGTTAATGTTTCTTATPCCGAGACAGCAAGAAC 1120
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
331 rThrGlyThr.....AsnSerHisIleGlySerThrAlaValArgL 345
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1121 CTGTTTATCATGCTCCAGTGGGTCAACAGTTATTCGCCACAGACTGAAT 1170
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 euAlaGlyAsnGluArgGlyAla.....Asn 353
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1171 AATGGAGAAATATTTCTCTTATTGACAAAGAAAGCTGAATTCATACT 1220
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 AsnGlyGlnAsnValThrPhe.....GluAsnAsnGlyThrLeuValLe 368
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1221 TACCAGCAACATCAACCAAGCGCGGCGTGTGTTTTCAGGGTAATT 1270
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
368 uAspGlnAsnIleAsnGlnGlyAlaGlyLeuPhePheLysGlyAspT 385
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1271 TTACGGTCTCGCTTAAACAAACGAA...ACGTGGCAAGCGCGCGCTT 1317
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
385 yrThrValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIle 401
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1318 CATATCAGTATGCGAGTACCGTTTACTTGGAAAGTAAACGCGTGGCAAA 1367
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 AspValAlaAspGlyLysValValTrpGlnValLysAsnProAsnGl 418
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1368 CGACCGCTGTCCAAAATCGCAAGCAGCGCTGCTGGTTCAAGCCAAAG 1417
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418 yAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrG 435
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1418 GGGAAACCAAGGCTCGGTGAGCGTGGCGCAGCGTAAAGTCACTTATGAT 1467
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 lyValAsnGlnGlyLysValGlyAspGlyThrValIleLeuAsn 451
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
1458 CAGCAGCGGACGATCAGGCAAGCAAAACACACCCCTTTAGTGAATCGGCTT 1517
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::  |||  ::
452 GlnGlnAlaAspAlaAspLysValGlnAlaPheSerGlnValGlyI 468
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
1518 GGTACAGCGGACGGGACGGTGCACCTGAATGCCGTAATCAGTTCACAC 1567
::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
468 eValSerGlyArgGlyThrLeuValLeuAsnSerProAspGlnIleAsn 485
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
1568 CCGCAAACTCTATTTCGGCTTTCGGCGCGACGCTTTGGATTTCAAGCGG 1617
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
485 roAsnAsnLeuTyPheGlyPheArgGlyArgLeuAspAlaAsnGly 501
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
1618 CATTCGCTTCGTCACCGCATTCAAATACCGATGAAGGGCGGATGAT 1667
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
502 AsnAspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgII 518
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
1668 TGTCAACACAAATCAAGACAAGAATCCACCGTTACCATTAACAGCAATA 1717
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
518 eValAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLys 535
|||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||||||  ::  |||  |||  ::  |||  ::
1718 AAGATATTACTACA 1731
::  |||  ::
535 erLeuIleThrAlaProGlnAsnLeuSerValTyGluIleArgAsnAsp 551
|||||  ::
1731 1731
552 TyrAspAspAspTyrTyGlyTyTyrSerTyArgLysProIlePr 568
|||||  ::
1731 1731
568 oGlnGlyLysAspLeuTyTyrLysAsnTyArgTyTyrAlaLeuLys 585
|||||  ::
1732 1732
585 erGlyGlySerValAsnAlaProMetProGluAsnGlyGlnThrGluAsn 601
|||||  ::
1741 RAC 1743
|||
602 AsnAspTrpIleLeuMetGlySerThrGlnGluAlaLysLysAsnAl 618
|||||  ::
1744 1744
618 aMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPheG 635
|||||  ::
1790 GCGAAGAGATGCAACCAACCAACGACGGCGGCTCAATCTGAATACCAA 1839
|||||  ::
635 lyGluGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsn 651
|||||  ::
1840 CCGGAAGAGCGGATCGCACTTTACTGCTTTCGGCGGACCAAAATTTAAA 1889
|||||  ::
652 GlyLysSerAlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAs 668
|||||  ::
1890 CCGCAATATCACGCAAAACAAACCGCAAACTGTTTTCACGGCGACAGCGA 1939
|||||  ::
668 nGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArgProt 685
|||||  ::
1940 CACCGCACCGCTTACATCATTAGGAACGGGTGGTCAAAAATGAAAGGT 1989
|||||  ::
685 hrProHisAlaArgAspPheValAsnLysSerSerAlaGlnLysAspAla 701
|||||  ::
1990 1990
702 HisPheSerLysAsnAsnGluValValPheGluAspAspTrpIleAsnAr 718
|||||  ::
2034 CACATTTAAACGGGAAACTTCCATATT 2080
|||||  ::
718 gThrPheLysAlaThrGluIleThrValAsnGlnSerAlaSerPheSerS 735
|||||  ::
2081 TTTCCCGCAATGTTGCCAAAGTGAAGCGGATTTGGCATTTTAAGCAATCAC 2130
|||||  ::
735 erGlyArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrAspAsn 751
|||||  ::
2131 GCCCAAGCAGTTTTCGGTGTGCGCCCGCATCAAGCCACACAATCTGTAC 2180
|||||  ::
```







[illegible]





seq\_name: pir2:I54632

seq\_documentation\_block:

tsh protein - Escherichia coli

C:Species: Escherichia coli

C>Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 08-Dec-2000

C:Accession: I54632

R:Provenance: D.L.; Curtiss, R.

Infect. Immun. 62, 1369-1380, 1994

A>Title: Isolation and characterization of a gene involved in hemagglutination by an avian

A:Reference number: I54632; MUID:94178945

A:Accession: I54632

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residue: 1-1377 <RES>

A:Cross-references: GB:I27423; NID:q469235; PIDN:AAA24698.1; PID:q469236

C:Superfamily: IGA-specific metalloendopeptidase

alignment\_scores:

Quality: 1125.50 Length: 1529

Ratio: 1.373 Gaps: 65

Percent Similarity: 53.630 Percent Identity: 26.226

alignment\_block:

US-09-303-518d-653 x I54632 ..

Align seq 1/1 to: I54632 from: 1 to: 1377

```

82  ITAGCCATATGCTGCTGCTCGCATTCGCCCAAGCCCGGGGGACA 131
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
42  ileProValLeuPheSerAlaGlySerLeu.....AlaGlyTh 54
132 GACTTATTTCGGCATCACTACCATCTATCGCGACTTTCGCCGAAATA 181
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
54  vAlaAsnAsnGluLeuGlyTyrGlnLeuPheArgaspPheAlaGluAsnL 71
182 TAGGCAAGTTTTCAGTCGCGGGCGAAGATATTGAGGTTTACACAAAAA 231
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
71  ysGlyMetPheArgProGlyAlaThrAsnIleAlaIleTyrAsnLysGln 87
232 GGGGAGTTGGTCGCAATTCATGATGACGAAAGCCCGCATGATTGTTTC 281
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
88  GlyGluPheValGly...ThrLeuAspLysAlaAlaMetProAspPheSe 103
282 TGTGTA...TCGCGTAACCGCGTGGCGCATTTGGCGGCGCATCAATATA 328
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
103  AlavalaspSerGluIleGlyValAlaThrLeuIleAsnProGlnTyrI 120
329 TGTGAGGTGGACATACGCGCGCTATACAAATGTTGATTGTTGGTGGC 378
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
120  leAlaSerValLysHisAsnGlyGlyTyrThrAsnValSerPheGly... 135
379 GAGGAAGCAATCCGATCAGCAGCCGCTTTCTTACCAAAATGTGAAGA 428
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
136  aspGlyGluAsn.....ArgTyrAsnIleValAspAr 146
429 AAATAATATAAAGCAGGACTTAACGGCCATCTTATGGCGCGCATATC 478
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
146  gAsnAsnAlaProSer.....LeuAspPheH 155
479 ATATCGCGGTTTGCAAAATTTGTCACAGATGCAAGAACCTGTGAGATG 528
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
155  IsAlaProArgLeuAspLysLeuValThrGluValAlaProThrAlaVal 171
529 ACCAGTTAT...ATGGATGGTGGAAATACCGTGATTTTAAATAATACCC 575
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
172  ThrAlaGlnGlyAlaValAlaGlyAlaTyrLeuAspLysGluArgTyrPr 186
576  TGCATGTTTGCATCGGAGCAGCAACAATATTGGCGGTCTGATGAAG 625
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
188  oValPheTyrArgLeuGlySerGlyThrGlnTyrIleLysAspSerAsnG 205

```

```

626  ACGAACCCCAATAACCGGAAAGTTTCATATCATATTCGAAGCGCATATCT 675
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
205  lyGlnLeuThr.....GlnMetGlyGlyAlaTyrSer 215
676  TGGCTCGTCGGTGGCAATACCTTTGCA.....CAAATGGA... 711
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
216  TripLeuThrGlyGlyThrValGlySerLeuSerSerTyrGlnAsnGlyG 232
712  .....TCAGGTGGTGGCACAGTCAACTTAGGTACGCAAAATAA 751
232  uMetIleSerThrSerSerGlyLeuVal.....PheA 243
752  AACATAGCCCATATGTTTTCACCAACAGAGAGCTCATTTGGCGACAT 801
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
243  spTyrLysLeuAsnGlyAlaMetProIleTyrGlyGluAlaGlyAspSer 259
802  GGCTCACCATGTTTATCTATGATGCCCAAAAGCAAAAGTGTGTTAATA 851
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
260  GlySerProLeuPheAlaPheAspThrValGlnAsnLysTrpValLeuVa 276
852  TGGGTATTGCAACAGCAACCCCTATATAGGAAAGCAATGGCTCC 901
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
276  lGlyValLeuThrAlaGlyAsnGlyAlaGlyGlyArgGlyAsnAsnTrpA 293
902  AGCTAGTTCGTAAGATGGTTCTATGATGAAATCTTCTGCTGGAGATCC 951
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
293  laValIleProLeuAspPheIleGlyGlnLysPheAsnGluAspAsnAsp 309
952  CATTACGATTCTACCAACCATCAAAATGGG.....AAATACTTTT 995
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
310  AlaProValThrPheArgThrSerGluGlyGlyAlaLeuGluTrpSerPh 326
996  TAACGACAATAATATGGCGCAGGAAATAATCGATGCAAAACATAACACT 1045
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
326  eAsn....SerSerThrGlyAlaGlyAlaLeuThrGlnGlyThrThrTh 342
1046  ATTCTCTACCTTATAGATTAAACACGACCGTTCATTTGTTAATGTT 1095
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
342  yrAlaMet..... 344
1096  TCATTATCCGAGACAGCAAGAACCTGTTTATCATGCTGCAGGTGGGT 1145
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
345  .....HisGlyGlnGlnGlyAs 350
1146  CAACAGTTATCGACCCAGACTGAATAATGGAGAAATATTTCTTTATTG 1195
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
350  nAsp.....LeuAsnAlaGlyLysAsnLeuIlePheGlnG 362
1196  ACAAGGAAAGTGATTTGATACCTTACCACACATCAACCAAGCGCG 1245
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
362  lyGln...AsnGlyGlnIleAsnLeuLysAspSerValSerGlnGlyAla 377
1246  GCGGTTTGTATTTTGAGGTAAATTTTACGGTCTCGCCTAAAAACAACGA 1295
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
378  GlySerLeuThrPheArgaspAsnTyrThrValThrThrSerAsnGlySe 394
1296  AACGTGGCAAGCGCGCGCTTCATATCATGATGATGACGATACCGTTACT 1345
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
394  rThrTrpThrGlyAlaGlyIleValValAspAsnGlyValSerValAsnT 411
1346  GGAAGTAACGCGGTGGCAACGACCGCTGTCCAAATTCGCAAGGC 1395
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
411  rpGlnValAsnGlyValLysGlyAspAsnLeuHisLysIleGlyGluGly 427
1396  ACCTGCTGTTTCAAGCCAAAGGGGAAACCAAGGCTCGCTCAGCGTGG 1445
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
428  ThrLeuThrValGlnGlyThrGlyIleAsnGluGlyGlyLeuLysValG 444
1446  CGACGGTAAAGTCATCTTAGATCAGCGGCGGAGATCAAGGCAAAAC 1495
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
444  yAspGlyLysValValIleuAsnGlnGlnAlaAspAsnLysGlyGlnValG 461

```

748 ..... ||| ||||::: ::: :||||::: ThrAlapheThrLeugluglYthrSerValAlaThr 760

2251 ..... TTAGCAAGACCGACATCAGAGGCAATGTGCAGCTTGGCGATCA 2294  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

761 LysAspAlaAspLysSerValPheasnGlyThrValasnLeuLeuAspAsnGI 777

2295 CGCTCATTTAAATCTCACAGGACTTGCCACACTCAACGCGCAATCTTAGT 2344  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::  
777 nSerValLeuAsnIleAsnAspIle.....PheasnGlyGlyIleGlnA 792

2345 CAGGCGGAGACAGCGCACTATACGGTTACGCCGAACGCCACCACAAC... 2391  
|| :||||::: |||  
792 Ia.....AsnAsnSerThrValAsnIle 799

2392 ...GGCAACCTCAGCCTCGTGGCGCAATGCCAAGCAACATTTAATCAAGC 2438  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

800 SerSerAspSerAlaValLeuGlyAsnSerThrLeuThrSerThrAlaLe 816

2439 CACATTAACGGCAACACATCGGCTTCGGACAAATGCTTCATTT... 2481  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

816 uAsnLeuAsnLysGlyAlaAsnAlaLeuAlaSerGlnSerPheValSera 833

2482 .....AATCTAACCAACACGCCGCTACAA... 2505  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

833 spGlyProValAsnIleSerAspAlaLaLeuSerLeuAsnSerArgPro 849

2506 .....AACSGCAGTCT 2516  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

850 AspGluValSerHisThrLeuLeuProValTyrrAspTyrrAlaGlySerTr 866

2517 GAGCGTT...TCCGACACGCTTAAGCAAGCAAGCTAAGC...CATTCGCGAC 2560  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

866 pAsnLeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrrSerMetL 883

2561 TCACGCGCAATGCTCCCCTAGCCGATAAG... 2589  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

883 euSerGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyGly 899

2590 .....GC 2591  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

900 GluGlyLeuLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrrSe 916

2592 AGTATTCCATTTTGAACACAGCGCTTTACCGGAAAAATCAGCGCGCGCA 2641  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

916 rLeuPheAsnGlyTyrrArgAsnIleTripsSerGlySerLeuAsn...Alap 932

2642 AGATACGCCANTACACTTAAAGACACGGAATGGACGCTGCGCTGGCG 2691  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

932 roAspAlaThrValSerMetThrAspThrGlnTripsrMetAsnGlyAsn 948

2692 ACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTCACCTCAATTC 2741  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

949 SerThrAlaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsn.. 964

2742 CGCCTATCGACACGATGCGGCGGCGCGCAACCGCAGATGCGCGCAGATG 2791

964 ..... 964

2792 CGCGCGCGCGGCTTCGCGCGCTTCCCTATTATCCGTTACGCGCCAAC 2841

964 ..... 964

2842 TCGCGAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAATTTGAACGG 2891

965 GlyGlyThrSerProPheThrThrLeuThrThrAspAsnLeuLeuAspAlaVa 981

2892 TCAGGGAACATTCGCTTTATGTCCGAACCTTCGCGCTACCGCAGCGGCA 2941  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

981 lGlnSerAlaPheValMetArgThrAspLeu.....AsnIlyAlaAspL 996

2942 AATTTGAAGCTGGCGGAAGATTCCGAAGGC...ACTTACACTTGGCTGTC 2988  
:||||::: :||||::: :||||::: :||||::: :||||::: :||||:::

996 ysLeuValIleAsnLysSerAlaThrGlyHisAspAsnSerIleTrpVal 1012  
2989 AAC.....AATACGGCAACAAACCC...GT 3011  
1013 AsnPheLeuLysLysProSerAsnLysAspThrLeuAspIleProLeuVa 1029  
3012 AAGTCTCGAGCAATGACG.....G 3031  
1029 lSerAlaProGluAlaThrAlaAspAsnLeuPheArgAlaSerThrArgV 1046  
3032 TAGTGAAGAAAGAACACACACCGCTGTCCGAAATCTTAATTCACCC 3081  
1046 alValGlyPheSerAspValThrProIleLeuSerValArgLysGluAsp 1062  
3082 CTGCAAAACGACGCTCGATCGCGCGCATGGCGTTATCAGCTTATCCG 3131  
1063 GlyLysLysGluTrpValLeuAspGly.....TyrGlnValAlaAr 1076  
3132 QAAAGACGGCGAG.....TTCGGCTGC 3154  
1076 GAsnAspGlyGlnGlyLysAlaAlaThrPheMetHisIleSerTyrA 1093  
3155 ATATCCGCTCAAGAA.....CAAGACTT 3180  
1093 snAsnPheIleThrGluValAsnAsnLeuAsnLysArgMetGlyAspLeu 1109  
3181 TCCGACAAATCGCGAAGCGGGGAAACACAGCGCGCTTGACGGCAAA 3230  
1110 ArgAspIleAsnGlyGluAlaGlyThrTrpValArgLeuLeuAsnGlySe 1126  
3231 ACAGGCA.....CAACTTG 3244  
1126 rGlySerAlaAspGlyGlyPheThrAspHisTyrThrLeuLeuGlnMetG 1143  
3245 CGCCAAACACAGCGGAAACACACACGCGCAAGCCTTGACGCGCTG 3294  
1143 lValAspArgLysHisGluLeuGlySerMetAspLeuPheThrGlyVal 1159  
3295 ATTGCGCGCGCGCAATGCCACGAAAGACAGAAAGTGTTCGCGAACCC 3344  
1160 MetAlaThrTyrThrAspThrAspAlaSerAlaAspLeuTyrSerGlyLy 1176  
3345 GCGCGCGAGCGCGCGGGAATGCGGCTATTCAGCGCGGAGGAAG 3394  
1176 sThrLysSerTrpGlyGly.....GlyPheTyrAlaSerGlyLeup 1190  
3395 AGAAACACGGTGCAGCGCGGATAAGACACCGCTTGGCGAAACACGCGC 3444  
1190 heArgSerGlyAlaTyrPheAsp.....ValIleAlaLysTyrIle 1203  
3445 GAAGCGGAAACCGCGCGCTTACCACCGCCTTCCCGCGCGCGCGCGC 3494  
1204 HisAsnGluAsnLys..... 1208  
3495 GCGCGGGATTGCGCAACCGACGACGCCAACCGCAACCCCAACCGCAGC 3544  
1209 .....T 1209  
3545 GCGACCTCAGCGCTTATGCCAATACGCGTTTGTAGTGAAATTTCCGCC 3594  
1209 rAspLeu.....AsnPheAlaGlyAlaGlyLysGlnAsnPheArgSer 1223  
3595 AGCTCAACAGCGTTTTCGCGGTACAGACGAA..... 3627  
1224 .....HisSerLeuTyrAlaGlyAlaGluValGlyTyrArgTyrHisLe 1238  
3628 TTGACCGCGCTGTTCGCGAAGACCGCGCAACCGCGCTTGTGCAACGCG 3676  
1238 uThrAspThrThrPheValGluProGlnAlaGluLeuValTrp..... 1252  
3677 GCATCCGGACACCAACACATACCGTTCGCAAGATTTCGCGCGCTACCGC 3726  
1253 .....GlyArgLeuGlnGlyGlnThrPheAsnTrpAsnAsp 1264

3727 CAACAAACCGACCTCGCCCAATCGTATGCAGAAAAACCTCGGCAGC... 3774  
1265 SerGlyMetAsp.....ValSerMetArgArgAsnSerValAsnPr 1278  
3775 .....GGGCGCGTCGCATCTCTTTTCGCACAAACCGGACCGGAACA 3817  
1278 oLeuValGlyArgThrGlyValVal.....SerGlyLysT 1290  
3818 CDTTC.....GACGACGGCATCGCACTCGGCACGGCTTCCGCCAC 3858  
1290 hrPheSerGlyLysAspTrpSerLeuThrAlaArgAlaGlyLeuHisTyr 1306  
3859 GGTGCGCTTTTCGGCAATACGCGATCGCGAGTTCGACATCGGCATCAG 3908  
1307 .....GluPheAspLeuThrAspSe 1313  
3909 CGCGGCGCGGGTTTGTAGTCGGCAGCCTTTCACAGCGCATCAGAGCA 3958  
1313 rAlaAspValHisLeuLysAspAlaAlaGlyGluHisGlnIleAsnGlyA 1330  
3959 AAATCGCGCGCGCTGCTGCAT...TACGGCATTCAGGCAAGATAC... 4002  
1330 rgLysAspSerArgMetLeuTyrGlyValGlyLeuAsnAlaArgPheGly 1346  
4003 .....CGCGAGGTTTCGCGCGGATT 4022  
1347 AspAsnThrArgLeuGlyLeuGluValGluArgSerAlaPheGlyLysTyr 1363  
4023 CGGCATCGAACCGCATCGCGCAACG...CGCTAT 4056  
1363 rAsnThrAspAlaIleAsnAlaAsnIleArgTyr 1375

seq\_name: pir2:S57664

seq\_documentation\_block:

IgA-specific metalloendopeptidase homolog sepA precursor - Shigella flexneri  
C:Species: Shigella flexneri  
C:Date: 19-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Dec-2000  
C:Accession: S57664; S69769; S69768  
R:Benjelloun-Touimi, Z.; Sansonetti, P.J.; Parsot, C.  
submitted to the EMBL Data Library, February 1995  
A:Description: Characterization of SepA, the major extracellular protein of Shigella  
A:Reference number: S57664  
A:Molecule type: DNA  
A:Residues: 1-1366 <BEN>  
A:Cross-references: EMBL:Z48219; NID:g886952; PIDN:CAA88252.1; PID:g886953  
R:Benjelloun-Touimi, Z.; Sansonetti, P.J.; Parsot, C.  
Mol. Microbiol. 17, 123-135, 1995  
A:Title: SepA, the major extracellular protein of Shigella flexneri: autonomous secre  
A:Reference number: S69768; MUID:96020667  
A:Accession: S69769  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-66;1077-1366 <BEW>  
A:Cross-references: EMBL:Z48219; NID:g886952; PIDN:CAA88252.1; PID:g886953  
A:Experimental source: strain M90T (serotype 5)  
A:Accession: S69768  
A:Molecule type: protein  
A:Residues: 57-61, Q, 63-72; X', 544-557; X', 1057-1068 <BEP>  
A:Note: 6-Glu, 14-Thr, 18-Gly and 26-Gln were also found  
C:Genetics:  
C:Gene: sepA  
C:Superfamily: IgA-specific metalloendopeptidase  
C:Keywords: extracellular protein  
F:1-36/Domain: signal sequence #status predicted <SIG>  
F:57-1366/Product: IgA-specific metalloendopeptidase homolog sepA #status predicted <

alignment\_scores:  
Quality: 996.00 Length: 1599  
Ratio: 1.233 Gaps: 64

Percent Similarity: 50.532 Percent Identity: 25.203

alignment\_block:

US-09-303-518D-653 x S57664

Align seg 1/1 to: S57664 from: 1 to: 1366

```
22 ACAACCGAAGACACACACCCGAAAGCCCTAAACCGCGCGCATCGCTTCTC 71
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
27 ThrCysLysSerHisArgLeuSerArg...ValIleLeuThrSe 42
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
72 GCCCGCTTACTGACCATATGCCGTGCTGCGCATTCGCCCAAGCCC 121
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
42 rValAlaLeuSerLeuSerSerAlaTrpProAlaLeuSerAlaThrV 59
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
122 GGGCGGACACACATATTTCGGCATCACTACCAATACATATCGGACTTT 171
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
59 alSerAlaGlu.....IleProTyrGlnIlePheArgASPhe 71
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
172 GCCGAAATAAAGCAAGTTGTCAGTCGGCGGCGAAGATATTGAGGTTTA 221
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
72 AlaGluAsnLysGlyGlnPheThrProGlyThrThrAsnIleSerIleTy 88
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
222 CAACAAAGAGGAGTGTGCGCAATCGATGACGAAAGCCCGCATGA 271
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
88 rAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysAlaPrometa 104
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
272 TTGATTTTCTGCTGCTATCGGTAAACGGCTGGCGGCAATTCGGGGCGAT 321
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
104 laAspPheSerSerAlaThrIleThrGlySerLeuProGlyAsp 120
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
322 .....CAATATATTGTGAGCGTGCACATACAGCGCG 353
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
121 HisThrLeuTyrSerProGlnTyrValThrAlaLysHisValSerGI 137
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
354 CTATAACAATGTTGATTTGTCGGGAGGAGCAATCCCGATCAGCACC 403
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
137 ySerAspThrMetSerPheGly..... 144
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
404 GCTTTTCTTACCAATTTGTGAAGAATAATTATAAGCA....GGACT 450
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
145 .....TyrAlaLysAsnThrTyrThrAlaValGlyThr 155
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
451 AACGGCCATCTTATGCGCGGATATCATATGCGCGTGTGACAAATT 500
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
156 AsnAsnSer...GlyLeuAspIleLysThrArgLeuSerLysLe 171
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
501 TGTACAGATGCAACACTGTTGAGATGACC..... 531
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
171 uValThrGluValAlaProAlaGluValSerAspIleGlyAlaValSerG 188
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
532 ..AGTTATGATGGTGGGAATACGCTGATTAAATAATACCCCTGAT 579
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
188 lAlaTyrGlnAlaGlyLysArgPheThrGluPhe..... 199
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
580 CGTGTTCGAATCGAGCAGGACAGCAATATTGGCGGCTGATGAAGACGA 629
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
200 ...TyrArgLeuGlyGlyMetGlnTyrVallys.....AspLys 212
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
630 ACCCAATACCGGAAAGTTCATATCATATTGCAAGCGCATATCTTGGC 679
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
212 sAsnGlyAsnArgThrGlnValTyr.....ThrAsnGlyGlyPheL 226
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
680 TCGTCGGTGGCAATACCTTTGCACAAATGATCAGGTGGTGC..... 723
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
226 euValGlyGlyThrValSerAlaLeuAsnSerTyrAsnAsnGlnMet 242
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
724 ...ACAGTCACCTTAGGTAGCAAAAAATTAAACATAGCCCATATGGTTT 770
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
243 IleThrAlaGlnThrGly...AspIlePheLeuIleProProAsnGlyPr 258
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
771 TTTTACCACAGGAGGCTCATTTGGCGAGTGGCTCACCATGTTTATCT 820
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
258 oLeuAlaAsnTyrLeuAsnMetGlyAspSerGlySerProLeuPheAlaT 275
821 ATCATGCCCAAGCAAAAGTGTAAATTAATGCGGTATTGCAACAGCG 870
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
275 yAspSerLeuGlnLysLysTrpValIleGlyValLeuSerSerGly 291
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
871 AACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTAGTTCGTAAAGATTG 920
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
292 ThrAsnTyr.....GlyAsnAsnTrpValValThrThrGlnAspPh 305
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
921 GTTCTATGATGAATCTTTGCTGGAGATACCATTCAGTATTCTACGAAC 970
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
305 eLeu..... 306
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
971 CACATCAAAATGGAAATACCTTTTAAAGCACAATAATAATGGCGCAGGA 1020
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
307 .....GlyGlnGlnProGlnAsnAsp..... 313
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1021 AAAATCGATGCCCAACATAAACACTATTCTCTACCTTATAGATTAATAAC 1070
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
313 ..... 313
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1071 ACGAACCGTTCAATGTTTAAATGTTTTCCTTATCCGAGACAGCAAGAAC 1120
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
314 .....PheAspLysThrIleAlaTyrThrSerGlyGluG 325
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1121 CTGTT.....TATCATGCTGCAGGTGGGTCAAC..... 1149
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
325 lYValLeuGlnTrpLysTyrAspAlaAlaAsnGlyThrGlyThrLeuThr 341
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1150 .....AGTTATCGACCCAGA.....CT 1166
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
342 GlnGlyAsnThrThrTrpGlyTyrAlaTrpLysGluArgLysLeuIleLe 358
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1167 GAATATGAGAAATATTTCTTATTGACAAAGAAAGTGAATTGA 1216
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
358 uAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGlyGluValV 374
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1217 TACTTACCAGCAACATCAACCAAGCGCGGCGGTTTGTATTTCGAGGT 1266
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 aLeuGlnAsnSerValAsnGlnGlyAlaGlyTyrLeuGlnPheAlaGly 390
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1267 AATTTTACGCTGCTCCAAATCGCAAGGACGCTGCTGTTCAAGCCAAA 1316
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
391 AspTyrArgValSerAlaLeuAsnGlyGlnThrTrpMetGlyGlyGlyI 407
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1317 TCATATCAGTATGCGCAGTACCGTTACTTGGAAAGTAAACGCGCGTGCNA 1366
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
407 eIleThrAspLysGlyThrHisValLeuTrpGlnValAsnGlyValAlaG 424
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1367 ACGACGCGCTGCCAAATCGCAAGGACGCTGCTGTTCAAGCCAAA 1416
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
424 lYAspAsnLeuHisLysThrGlyGluGlyThrLeuThrValAsnGlyThr 440
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1417 GGGGAAACCAAGGCTCGCTGCGTGGCGGACGCTGCTGTTGATTTCGAGCG 1466
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
441 GlyValAsnAlaGlyLysValGlyAspGlyThrValIleLeuAs 457
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1467 TCAGCAGCGGAGCATCAAGGCAAAACCAACGCTTGTAGTAATCGGCT 1516
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
457 nGlnGlnAlaAspArgAspGlyLysValGlnAlaPheSerSerValGlyI 474
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1517 TGGTCAGCGGAGGGGACGCTGCAACTCAATGCCGATATCAGTTCAC 1566
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
474 lAlaSerGlyArgProThrValValLeuSerAspSerGlnGlnValAsn 490
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1567 CCGGCAAACTCTATTTCGCTTTCGCGCGGACGCTTGTGATTTCGAGCG 1616
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
491 ProAspAsnIleSerTrpGlyTyrArgGlyGlyArgLeuGluLeuAsnG 507
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1617 GCATTGCTTTCGTTCCACCGCATTAATAATACCGATGAAGGGCGATGA 1666
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
507 yAsnAsnLeuThrPheThrArgLeuGlnAlaAlaAspTyrGlyAlaIleI 524
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```



1016 ..... 1016  
3259 GCGGAAAGACAAACGCGCAAGCCTTACGCGCTGATTTCGGCCGCGG 3308  
1017 .....LeuIleThrAlaProly 1022  
3309 CAATGCCACCGAAAGACGAGAAAGTGTTCGGAACCGCGCCGCGCAGCAG 3358  
1022 sGlySerAspGluLysThrPheThr.....AlaG 1032  
3359 GCGGGGAAATCCCGCATTTATGACGCGGAGGAGAGAAAAACGGGTG 3408  
1032 lyThrGlnGlnIleGlyPheSerAsnValThrProValIleSerThrGlu 1048  
3409 CAGCGGATAAAGACACCGCTTGGCGGAACACGCGGAACCGCAACCCG 3458  
1049 LysThrAspAlaThr.....LysTrpValLeuThrGI 1060  
3459 GCGGCTACACCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGC 3508  
1060 yThrGlnThrThraAlaAspAlaGlyAlaSerLysAlaAlaLysAsp.... 1075  
3509 CGCAACCGCAGCCCAACCCACCCCAACCGCAGCGGACCTGTATCAGC 3558  
1075 ..... 1075  
3559 CGTTATGCCAATAGCGGTTTGTAGTGAATTTTCGCCACGCTCAACAGCGT 3608  
1076 ...PheMetAlaSerGlyTyrLysSerPheLeuThrGluValAsnAsnLe 1091  
3609 TTTCGCGGTACAGACGAAATTCGACGCGGTGTTTTCGCGAAGACCGCGCA 3658  
1091 uAsnLysArgMetGlyAspLeu.....ArgAspThrGlnG 1103  
3659 AC.....GCCCTTGGACAGCGGATTCGCGGACACAACTACCGT 3702  
1103 lyAspAlaGlyValThrPalaArgIleMetAsnGlyThr.....Gly 1116  
3703 TCAGCAAGATTTCCGCGCTACCGCAACAAACCGCACTGCGCAAAATCGG 3752  
1117 SerAlaAsp...GlyAspTyrSerAspAsnTyrThrHisValGlnIleGI 1132  
3753 TATGCAGAAA.....AACCTCGGAGCGGCGCGG 3781  
1132 yValAspArgLysHisGluLeuAspGlyValAspLeuPheThrGlyAlaL 1149  
3782 TCAGCATCTGTGTTTCGCAACACCGGACCGGAACACCTC..... 3822  
1149 eLeuThrTyrThrAspSerAsnAlaSerSerHisAlaPheSerGlyLys 1165  
3823 GACGACGGCATCGGCAACTCGGCAACGCGCTTGCACCGGTTGCCGTTTCGG 3872  
1166 AsnLysSerValGlyGly..... 1171  
3873 CAATACGGCATCGGCAAGTTTCACATCGGCATCGCGCGCGCGGTT 3922  
1172 .....GlyLeuTyrAlaSerAlaLeup 1179  
3923 TTATAGCGGCGCTTTCAGACGGCATCAGAGGCAAAATCGCCCGCC 3972  
1179 HeAsnSerGlyAlaTyrPheAspLeuIle...GlyLys..... 1190  
3973 GTCTGCATTACGGCATTCAGCAAGATACCGCGCAGGTTTCGGCGGATT 4022  
1191 TyrLeuHisHisAspAsnGln.....HisThrAlaAsnPheAlaSerLe 1205  
4023 CGGCATCGAACCCCATCGGCGCAACGCGCTATTTCGTCCAAAAGCGG 4072  
1205 uGlyThrLysAspTyrSerHisSerHisSerTrpTyrAlaGlyAlaGluValG 1222  
4073 ATTACCGATAC.....GAA 4086  
|||||

1222 lyTyrArgTyrHisLeuThrLysGluSerTrpValGluProGlnIleGlu 1238  
4087 ACGTCAATATCCGACCCCGGCTTGCATTCAACCGCTACCGCGCGG 4136  
1239 LeuValTyrGlySerValSerGlyLysAlaPheSerTrpGluAlaArgI 1255  
4137 CATTAAGGCGAGATTATTCAACCGCGCAACACATTTCCATCACGC 4186  
1255 yMetAla.....LeuSerMetLysAspLysAspTyrAsnProLeuIleG 1270  
4187 CTATTGAGCCTGCTTATACCGGATGCGCTTCGCGC.....AAA 4227  
1270 lyArgThrGlyValAspValGlyArgAlaPheSerGlyAspAspTrpLys 1286  
4228 GTCCCAACCGCG.....GTCAATAC 4247  
1287 IleThrAlaArgAlaGlyLeuGlyTyrGlnPheAspLeuLeuAlaAsnGI 1303  
4248 CGCGGTATTGCGCGAGGATTTCGCGCAACCGCGAGTGGCGAATGG.... 4293  
1303 yGluThrValLeuGlnAspAlaSerGlyGluLysArgPheGluGlyGluL 1320  
4294 .....GGCGTAAACCGCGCAATCAAA 4314  
1320 ysAspSerArgMetLeuMetThrValGlyMetAsnAlaGluIleLys 1335

seq\_name: pir2:S61315

seq\_documentation\_block:

Iga-specific metalloendopeptidase (EC 3.4.24.13) - *Neisseria meningitidis* (strain EN3)  
N:Alternate names: Iga proteinase; immunoglobulin A1 proteinase  
C:Species: *Neisseria meningitidis*  
A:Variety: strain EN3771; strain ETH2  
C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
C:Accession: S61315; S61316  
R:Lomholt, H.; Poulsen, K.; Mogensen, K.  
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria meningitidis* strains EN3771 and ETH2  
A:Reference number: S61314; MUID:95302961  
A:Accession: S61315  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-496 <LOM>  
A:Cross-references: EMBL:X82468; NID:g732865; PIDN:CAA57851.1; PID:g732866  
A:Experimental source: strain EN3771  
A:Accession: S61316  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-496 <LOL>  
A:Cross-references: EMBL:X82469; NID:g732887; PIDN:CAA57852.1; PID:g732888  
A:Experimental source: strain ETH2  
A:Genetics:  
A:Gene: iga  
C:Superfamily: Iga-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

alignment\_scores:  
Quality: 883.50 Length: 540  
Ratio: 2.489 Gaps: 19  
Percent Similarity: 65.741 Percent Identity: 39.074

alignment\_block:

US-09-303-518D-653 x S61315 ..

Align seg 1/1 to: S61315 from: 1 to: 496

202 CGAAAGATATTGAGTTTACAAACAAAGGAGTGTGGTGGCAATC 251  
||| |||.....||| |||.....  
1 AlathrAspLeuSerValLysAsnLysGlnGlnAsnIleGlyAsnAl 17  
252 GATGACGAAGCCCGATGATTGATTTTCTGTGTATCGCGTAAC...G 298  
:.....:|||||.....: |||

17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34  
299 CGGTGGCGCATTTGGCGGCGATCAATATATGTGAGCGTGGCACATAAC 348  
34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50  
349 GCGGCG.....TATAACAATGTTGATTTGGTGGCGAGGG 383  
51 LysGlyAspGluIleSerTyrTyrGlyHisAsnGlyHisLeuAspVa 67  
384 AAGCAATCCGATCAGCACCGCTTTCTTACCAAAATTTGTGAAGAATA 433  
67 lSerAsnAspGluAsn.....GluTyrArgSerValAlaGlnAsnA 81  
434 ATTATAAGCAGGAGGACTAACGCCATCTTATGGCGG..... 471  
81 spTyrGluProAsnLysAsnTrpHis....HisGlyAsnGlnGlyArgLeu 96  
472 ..GATTATCATATCCCGCTTTCACAAATTTGTACAGATGCAAGAAC 518  
97 GluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPr 113  
519 TCTTCAGATGACCGAGTTATATGATGGTGGAA.....TACCTCATTTAA 565  
113 oIleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysA 130  
566 ATAAATACCCGTGATCGTGTTCGAATCGGAGCAGCAGCAATAATATGGCG 615  
130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 144  
616 TCTGATGAAGCAGCAACCAATACCGGAAAGTTTCATATCATATATGCAAG 665  
145 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157  
666 CCATATCTTGGCTGCGTGGTGGCAATACCTTT.....G 700  
157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174  
701 CACAAATGGATGATGAGTGGTGGCAGAGTCACTTAGGTAGCGAGAAATTT 750  
174 hrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLys 190  
751 AAACATAGCCCA.....TATGGTTTTTTTACC 776  
191 HisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuTh 207  
777 ACAGAGGCTCATTTGGCGAGCGTCCACCAATGTTTATCATATGATG 826  
207 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspL 224  
827 CCCAAAGCAAAAGTGTAAATTAATGGGGTATTGCAACAGCAACCCC 876  
224 ysGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTriaGly 240  
877 TATATAGGAAAGCAATGGCTTCAGCTAGTTCGTAAGATTGG....TT 923  
241 Tyr.....GlnLysAsnSerTrpGln.....GluTrpAsnIle 251  
924 CTATGATGAATCTTGTGCTGGAGATACCCATTCAGTATTCTACGAACCC 973  
251 eTyrLysLysGluPheAlaAspGluIleLysGln..... 262  
974 ATCAAAATGGGAATACTTTTTTAACGACAATAATAATGGCGCAGAAAA 1023  
263 .....ArgAspAsn.....AlaGlyThr 268  
1024 ATCGATGCCAAACATAAACACATCTTCTACCTTATAGATTAATAAACACG 1073  
269 IleLysGlyTyrGlyGluHis.....TrpLysThr 280  
1074 AACCGGTCAATGTTTAAATGTTCTTTTATCCGAGACAGACAGAACCTG 1123  
280 rGlyThr.....AsnSerHisIleGlySerThrAlaValArgLeuA 294

1124 TTTATCATCTGCAGGTGGGTCAACAGTTATCGACCCAGACTGAATAT 1173  
294 laGlyAsnGluArgGlyAla.....AsnAsn 302  
1174 GGAAGAAATATTCCTTTATTACAAAGAAAGGTGAATTGATCTTAC 1223  
303 GlyGlnAsnValThrPheGluAsp.....AsnGlyThrLeuValLeuAs 317  
1224 CAGCAACATCAACCAAGCGCGGGTGTATTTTGTAGGGTAAATTTTA 1273  
317 pGlnAsnIleAsnGlnGlyAlaGlyLeuPhePheLysGlyAspTyrT 334  
1274 CGTCTCTCGCTTAAACAAACGAA...ACGTGGCAAGCGCGGGCTTCAT 1320  
334 hrValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyLeuAsp 350  
1321 ATCAGTGATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAACGA 1370  
351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367  
1371 CGCCTGTGCAAAATCGCAAGGACGCTGCTTCAAGCCAAAGGGG 1420  
367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 384  
1421 AAACCAAGCTCGTCAGCGTGGCGCGGTAAAGTCATCTTATAGATCAG 1470  
384 alAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400  
1471 CAGCGCGAGCATCAAGGCAAAACAAAGCTTTAGTGAATCGGCTGTGT 1520  
401 GlnAlaAspAlaAspLysValGlnAlaPheSerGlnValGlyIleVa 417  
1521 CAGCGCGAGGAGCGTGCACACTGAATGCCATATCACTTCAACCCCG 1570  
417 lSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProA 434  
1571 ACAACTCTATTTTCGCGTTCGCGCGGCGGTGGATTGATTTGAACGGCAT 1620  
434 spAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450  
1621 TCCTTTCGTTCCACCGCATTCAAATACCGATGAAGGGCGGATGATGTTGT 1670  
451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467  
1671 CACACCAATCAAGCAAAAGATCCACCGTACCATTACAGCAATAAAG 1720  
467 lAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLysSerL 484  
1721 ATATTACTACACCGCAAT 1740  
484 euleThrAlaProGlnAsn 490

seq\_name: pir2:S61319

seq\_documentation\_block:  
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr  
C:Species: Neisseria meningitidis  
A:Variety: NG117  
C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
R:Accession: S61319  
R:Lombolt, H.; Poulsen, K.; Mogens, K.  
Mol. Microbiol. 15, 495-506, 1995  
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss  
A:Reference number: S61314; MUID:95302961  
A:Accession: S61319  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-508 <DOM>  
A:Cross-references: EMBL:X82483; NID:g732881; PIDN:CAA57866.1; PID:g732882  
C:Superfamily: Iga-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase











974 ATCAAAATGGAAATACCTTTTAAACGACATAATAATGCGCAGGAAA 1023  
263 .....ArgAspAsn.....AlaGlyThr 268  
1024 ATCCATGCCAAACATAACACTATCTCTACCTATAGATTAACACACG 1073  
269 IleLysGlyTyrGlyGluHisHis.....TrpLysThr 280  
1074 AACCGTTCATATGTTTAAATGTTCTTTATCCGACAGCAGACGACCTG 1123  
280 rGlyThr.....AsnSerHisIleGlySerThrAlaValArgLeuA 294  
1124 TTTATCATGCTGCAGGTGGGCTCAACAGTTATCCACCCAGACTCAATAAT 1173  
294 laGlyAsnGluArgGlyAla.....AsnAsn 302  
1174 GGAGAAATATTCCTTTATGACAAAGGAAAGGTCAATGATCTTAC 1223  
303 GlyGlnAsnValThrPhe.....GluAsnAsnGlyThrLeuValLeuAs 317  
1224 CAGCAACATCAACCAAGCGCGGCTTGTATTTGAGGTAATTTA 1273  
317 pGlnAsnIleAsnGlnGlyAlaGlyLeuPheLysGlyAspTyr 334  
1274 CGGTCTCGCTTAAACCAACGAA...ACGTGCAAGCGCGGCTTCAT 1320  
334 hrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAsp 350  
1321 ATCAGTGTATGCGAGTACCTTACTTGAAGTAACGGCGTGCAACGA 1370  
351 ValAlaAspGlyLysValValTrpGlnValLysAsnProAsnGlyAs 367  
1371 CCGCTCTCCAAATCGGCAAGCGCTGCTGTTCAAGCAAGGAGG 1420  
367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 384  
1421 AAACCAAGCTCGTCAGCGTGGCGGCGGTAAGTCACTCTAGATCAG 1470  
384 aIAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400  
1471 CAGCGCGAGCATCAAGGCAAAACAAAGCCTTAGTGAATCGGTTGGT 1520  
401 LysAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleVa 417  
1521 CAGCGCAGGGGAGCGGTGCAACTGAATGCCGATATCAATCAACCCCG 1570  
417 lSerGlyArgArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProA 434  
1571 ACAACTCTATTCGCTTCGCGGCGGAGCTTGGATTGACGGGCAT 1620  
434 sPasnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 1621  
1621 TCGCTTTCGCTCCACCGCATTCAAAATCCGATGAAGGGCGCATCTCT 1670  
451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467  
1671 CAACCAATCAAGCAAGAATCCACCGTTCACCATACAGGCAATAAG 1720  
467 lAsnHisAsnThrGlyHisThrSerThrIleThrLeuThrGlyLysSerL 484  
1721 ATATTACTACACCGCGCAATACAACTTG 1752  
484 eulleThr.....AsnProAsnSerLeu 491

seq\_name: pir2:S61329

seq\_documentation\_block:

Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fragm

C:Species: Neisseria meningitidis

A:Variety: NGC16

C:Date: 23-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000

C:Accession: S61329

R:Lomholt, H.; Poulsen, K.; Mogensen, K.

Mol. Microbiol. 15, 495-506, 1995

A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria meningitidis

A:Reference number: S61314; MUID:95302961

A:Accession: S61329

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-508 &lt;LOW&gt;

A:Cross-references: EMBL:X82485; NID:g732869; PIDN:CAA57868.1; PID:g732870

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

alignment\_scores:

Quality: 859.50 Length: 547

Ratio: 2.463 Gaps: 18

Percent Similarity: 63.803 Percent Identity: 38.391

alignment\_block:

US-09-303-518D-653 x S61329 ..

Align seg 1/1 to: S61329 from: 1 to: 508

202 GCGAAGATATTGAGCTTTACACAAAAGGGAGTTGGTCGCAATC 251

1 AlaThrAspLeuSerValLysAsnGlnGlyGlnAsnIleGlyAsnAl 17

252 GATGACGAAAGCCCGATGATGATTTTCTGTGATCGGTAC...G 298

17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

299 CGGTGCGGCGCATTCGCGGCGCATCAATATATTGTGCGGTGCACAT... 345

34 rGileAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

346 .....AACGGCGCTATAACAATGTTGATTT 371

51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

372 TGTGCGGAGGAGCAATCCGATCAGCACCGCTTTTCTTACCAATG 421

67 pValAlaAspLysGluAsn.....GluTyrArgValV 78

422 TGAAGAATAATATAAACAGGAGCTAACGCGCATCTTATGCGGCG 471

78 aGluGlnAsnAsnTyrGluPro.....HisLysAlaTrpGly 90

472 .....GATTATCATATGCGCGCTTTGCAAA 497

91 AlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLy 107

498 ATTTGTACACATGCAGACCTGTGTAGATGACACAGTTATGATGGGT 547

107 spHeValThrGluValAlaProIleAlaProThrAspAlaGlyGlyL 124

548 GG...AAATACGCTGATTAAATAATACCTGATCGTTTCGAATCGGA 594

124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140

595 GCAGGCGACAA.....TATTGGCGTCTGTGATGAAGACGACCAATAA 638

141 AlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAs 157

639 CCGGAAAGTTCATAT.....CATATTGCAAGCGCATATCTTGGC 679

157 n...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrA 173

680 TCGTCGCTGGC.....AAT 693

173 lAlaAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsn 189

694 ACCTTCCACAAATGATCAGGTGGTGGCAGCTCACTTAGGTAGCGGA 743

190 ThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaG1 206  
744 AAAAATTAACATAGC...CCATATGGTTTATACCAACAGGAGCTCAT 790  
206 uGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyVal1 223  
791 TTGGCGACAGTGGCTCACCAATGTTATCTATGATGCCCAACAAAG 840  
223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239  
841 TGGTAAATTAATGGGTATTCGAACAGGCAACCCCTATATAGGAAGAAG 890  
240 TrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLysLys 256  
891 CAATGGCTTCAGCTAGTTCGTAAGATGTTCTATCATCAATCAATCTTTG 940  
256 TrpGlnGluTrpAsnIleTyrLysLysGluPheAlaAspLysLys 273  
941 CTGGAGATACCCATTTCAGTATTCTACGAACACATCAATGGAATATAC 990  
273 InArgAsp..... 275  
991 TTTTAAACGACATAAATAATGGCCGAGAAATAATCGATGCCAACAATAA 1040  
276 .....AsnAlaGlyThrIleLysGlyAsnGlyGluHis.. 286  
1041 ACACATATCTCTACCTTATAGATTAAACACACGACCGCTTCAATGTTTA 1090  
287 HisTrpAsnIleThrPheGlyThr.....A 295  
1091 ATGTTCTTTTATCCGAGACAGCAAGAACCTGTTTATCATGCTGCAGGT 1140  
295 snSerHisIleGlySerThrAla.....ValArgLeuAlaGly 307  
1141 GGGGTCAACAGTATTCGACCCAGACTGAATATGAGAAATATTTCTCTT 1190  
308 AsnGluArgAsp.....AlaAsnAsnGlyGlnAsnValThrPh 320  
1191 TATTGCAAAAGCAAGGTGAATTCATCTTACACAGCAACATCAACCAAG 1240  
320 eGluAsp.....AsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnG 335  
1241 GCGCGGCGGTTTGTATTTAGGTAATTTTACGGTCTCGCCTAAAAAC 1290  
335 lYalaglyGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsn 351  
1291 ACGAA...ACGTGGCAAGCGCGGCTTCATATCATGATGGCAGTAC 1337  
352 AsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLys 368  
1338 CGTTACTTGAAGTAAACGGCTGGCAACACGACCGCTGTCCAAATCG 1387  
368 sValValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleG 385  
1388 GCAAGGACCGCTGCTGTTCAAGCAAAAGGGGAAACCAAGGCTCGGTC 1437  
385 lYlsglyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeu 401  
1438 ACGGTGGGCGAGGTAAAGTCAATCTAGATACAGCGGCGGAGCATCAAG 1487  
402 LysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspSerAsnG1 418  
1488 CAAAACCAAGCCTTTAGTGAATCGGTTGTCAGCGCGAGGGGAGCGG 1537  
418 nLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyThrL 435  
1538 TGAACCTGAATCCGATATATCAAGTCAACCCGACAACTATTTTCGGC 1587  
435 euValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyrPheGly 451  
1588 TTTTCGGCGGAGGTTTGGATTGACCGGCAATTCGCTTCCTCCACCG 1637  
452 PheArgGlyGlyArgLeuAlaAsnGlyAsnAspLeuThrPheGluH1 468

1638 CATTCAAAATACCGATGAAGGGCGATGTTCTCAACCAATCAAGACA 1687  
468 sIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrGlyH 485  
1688 AGAATCCACCGTTACCATACAGGCAATAAGATATTACT 1728  
485 isAlaSerThrIleThrLeuThrGlyLysSerIleuThr 498

seq\_name: pir2:S61322

seq\_documentation\_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr  
C:Species: Neisseria meningitidis

A:Variety: HF54

C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000

C:Accession: S61322

R:omolc, H.; Poulsen, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A:Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss

A:Reference number: S61314; MUID:95302961

A:Accession: S61322

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-496 <LOW>

A:Cross-references: EMBL:X82473; NID:g732877; PIDN:CAA57856.1; PID:g732878

C:Superfamily: IgA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

alignment\_scores:

Quality: 858.50 Length: 540

Ratio: 2.432 Gaps: 19

Percent Similarity: 65.370 Percent Identity: 38.333

alignment\_block:

US-09-303-518D-653 x S61322 ..

Align seg 1/1 to: S61322 from: 1 to: 496

202 GCAGAGATATTGAGTTTACACAAAAAGGGAGTTGTCGGCAATC 251

1 AlathrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

252 GATGACGAAAGCCCGCATGATGATTTTCTGCTGATCGCGTAAC...G 298

17 aLeuSerAsnValPrometileAspPheSerValAlaAspValAsnArgA 34

299 GCGTGGCGGATTCGGCGGCGATCAATATATTGACCGTGGCACAATAC 348

34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

349 GCGCGC.....TATACAATGTTGATTTTGGTCGGGAGGG 383

51 LysGlyAspGluIleSerTyrTyrGlyHisAsnGlyHisLeuAspVa 67

384 AAGCAATCCGATCAGCACCGCTTTCTTACCAAAATGTGAAAGAAATA 433

67 lSerAsnAspGluAsn.....GluTyrArgSerValAlaGlnAsnA 81

434 ATTATAAGCAGGACATAAGCGCCATCTTATGCGGC..... 471

81 sptyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeu 96

472 ...GATTATCATATGCGCGTTCACACAAATTTGTACAGATGCAGAAC 518

97 GluAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPr 113

519 TGTTCAGATGACCACTTATATGATGGTGGTGA...TACGCTGATTAA 565

113 oileAlaProThrSerAlaGlyGlyValGluThrTyrLysAspLysA 130

566 ATAAATACCTGATCGTTCGAATCGGACGACGACACATATTGCGCG 615

```

130 snArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 144
166 TCTGATGAAGACGACCAATAACCGGAAAGTTTCATATCATATTCGAAG 665
145 .....GluTyrAsnSerArgTyrAsnMetThrGluLeuSerAr 157
666 CGCATATCTTGCTCGTGGTGGCGAATACCTTT.....G 700
157 gAlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValT 174
701 CACAAATATGATCAGGTGGGCGACAGTCAACATAGTAGCGCAAAAT 750
174 hrSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLys 190
751 AAACATAGCCCA.....TATGGTTTTTTTAAOC 776
191 HisHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuTh 207
777 AACGAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCATGATG 826
207 rAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspL 224
827 CCAAAAGCAAAAGTGGTAAATTAATGGGTATTCGAAACAGCAACCC 876
224 ySgInGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGly 240
877 TATATAGGAAAGCAATGGCTCCAGTAGTTCGTAAAGATTGG...TT 923
241 Tyr.....GlnLysAsnSerTrpGln.....GluTrpAsnIle 251
924 CTATGATGAATCTTGGTGGGATACCATTCAGTATTCAGTACCAAC 973
251 eTyLysLysGluPheAlaAspGluLysGln..... 262
974 ATCAAAATGGAAATACTTTTTTAACGACAATAAATGGCGCAGGAAAA 1023
263 .....ArgAspAsn.....AlaGlyThr 268
1024 ATCGATGCCAAACATAAACACTATTCCTACCTTATAGATTAAACACG 1073
269 IleLysGlyTyrGlyGluHis.....TrpLysThrTh 280
1074 AACCGTTCAATGTTTAAATGTTTCTTATTCGAGACAGCAAGAACCTG 1123
280 rGlyThr.....AsnSerHisIleGlySerThrAlaValArgLeuA 294
1124 TTATCATCTCGAGTGGGGTCAACAGTTATCGACCCGACTGAATAAT 1173
294 laGlyAsnGluArgGlyAla.....AsnAsn 302
1174 GGAGAAATATTTCCTTTATTCACAAAGAAAGGTGAATGATCTTAC 1223
303 GlyGlnAsnValThrPhe.....GluAsnAsnGlyThrLeuValLeuAs 317
1224 CAGCAACATCAACCAAGCGCGGTGTTGATTTGAGGTAAATTTA 1273
317 pGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPheLysGlyAspTyrT 334
1274 CGGTCTCGCTTAAACACACGAA...ACGTGGCAAGCGCGCGCTTCAT 1320
334 hrValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAsp 350
1321 ATCAGTGTGCGAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAACGA 1370
351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367
1371 CGGCTGTCCAAATCGCAAGGACCGCTGCTGTTCAACCCAAAGGG 1420
367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGly 384
1421 AAACCAAGGCTCGCTGAGCGTGGCGACGCTAAAGTTCATCTAGATCAG 1470

```

```

384 aLasnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400
1471 CAGCGGAGCATCAGGCAAAACAAAGCCTTTAGTAAATCGCGTGGT 1520
401 GlnAlaAspAlaAspLysValGlnAlaPheSerGlnValGlyIleVa 417
1521 CAGCGGAGGCGGCGTCCAACTGAATGCCGATAATCAGTTCAACCCG 1570
417 lSerGlyArgArgThrLeuValLeuAsnSerProAspGlnIleAsnProA 434
1571 ACAAACTCTATTTCCGCTTTCGCGCGGACGTTTGGATTGAACGGCAT 1620
434 snAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 450
1621 TCGCTTTTCCTCCACCGCATTCAAATACCGATGAAGGGCGGATCATGT 1670
451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467
1671 CAACCAATCAAGACAAGAAATCCACGTTTACCATTACAGGCAATAAG 1720
467 lAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLysSerL 484
1721 ATATTACTACAACCGCAAT 1740
484 euIleThrAlaProGlnAsn 490
seq_name: pir2:S61327
seq_documentation block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis (fr
A:Species: Neisseria meningitidis
A:Variety: NCG80
C:Date: 23-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61327
R:Lombolt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961
A:Accession: S61327
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown
A:Residues: 1-496 <LOM>
C:Cross-references: EMBL:X82479; NID:g732883; PIDN:CAA57862.1; PID:g732884
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

```

```

alignment_scores:
Quality: 856.50 Length: 536
Ratio: 2.433 Gaps: 20
Percent Similarity: 65.672 Percent Identity: 38.806
alignment_block:
US-09-303-518D-653 x S61327 ..
Align seg 1/1 to: S61327 from: 1 to: 496
202 GCGAAAGATATTGAGTTTACACAAAGGAGGAGTTGGTCGCAATC 251
1 AlathrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17
252 GATGACGAAAGCCCGGATGATTTCTGTTATCGCGCTAAC...G 298
17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34
299 CGGTGGCGGATTCGCGGCGGATCATATATTTGAGCGTGGCACAATAAC 348
34 rgrThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50
349 GCGGCG...TATACAAATGTTGATTTTGGTGGGAGGG 383
51 LysGlyAspGluIleSerTyrTyrGlyHisAsnGlyHisLeuAspVa 67

```



317 pGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrT 334  
 1274 CGGTCTCGCCTAAAAACAACGAA...ACGTGGCAAGCGCGGGCGTTCAT 1320  
 334 hrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAsp 350  
 1321 ATCAGTGATGGCAGTACCGTTTACTTGGAAAGTAACGCGGTGGCAACGA 1370  
 351 ValAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 367  
 1371 CCGGCTGTCAAAATCGGCAAGGCACGCTGCTGTTCAAGGCAAAAGGG 1420  
 367 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 384  
 1421 AAAACCAAGGCTCGGTGACGCTGGCGGACGGTAAAGTCAATCTTAGATCAG 1470  
 384 aAlAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 400  
 1471 CAGCGCAGCATCAAGCAAAAAACAACGCTTTAGTGAATCGGCTGGTGT 1520  
 401 GlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleVa 417  
 1521 CAGCGCAGGGGACGGGTGCAACTGAATGCCGATAAATCAGTTCAACCCGG 1570  
 417 lSerGlyArgGthrLeuValLeuAsnSerSerAsnGlnIleAsnProA 434  
 1571 ACAAACTATTTCGGCTTCGCGCGGACGTTTCGATTGTGAACGGCAT 1620  
 434 spAsnLeuTyrPheGlyPheArgGlyArgLeuAspAlaAsnGlnGlyAsn 450  
 1621 TCGCTTTCTCCACCGCATTCAAAATACCGATGAAGGGCGGATGTTGT 1670  
 451 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 467  
 1671 CAACCAAACTCAACACAAGAATCCACGGTACCATTACAGGCAATAAAG 1720  
 467 lAsnHisAsnThrGlyHisAlaSerThrIleThrLeuThrGlyLysSerL 484  
 1721 ATATTACT 1728  
 484 euIlleThr 486

seq name: pir2:S61320

seq\_documentation\_block:  
IgaA-specific metalloendopeptidase (EC 3.4.24.13) homolog - *Neisseria meningitidis* (fr  
C:Species: *Neisseria meningitidis*  
A:Variety: NCC65  
C:Date: 20-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
C:Accession: S61320  
R:Domholt, H.; Poulsen, K.; Mogens, K.  
Mol. Microbiol. 15, 495-506, 1995  
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neiss*  
A:Reference number: S61314; MUID:95302961  
A:Accession: S61320  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-508 <LOM>  
A:Cross-references: EMBL:X82484; NID:g732871; PIDN:CAA57867.1; PID:g732872  
C:Superfamily: IgaA-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

```

alignment_scores:
  Quality: 851.50      Length: 553
  Ratio: 2.426         Gaps: 18
  Percent Similarity: 63.472  Percent Identity: 37.794

alignment_block:
  US-09-303-518D-653 x S61320  ..
  Align seq 1/1 to: S61320 from: 1 to: 508

```

985 AAATACTTTTTTAAACGACAATAATAATGGCGAGGAAAAATCGATGCCAA 1034  
 276 .....AsnAlaGlyThrIleLysGlyAsnGlyG 285  
 1035 ACATAACACATATCTCTACCTTACGATTAAAAACACGAACCGTTCAT 1084  
 285 uHis...HisTrpAsnIleThrPheGlyThr..... 294  
 1085 TGTTTAAATGTTCTTTATTCGACGACACAGAGAACCTGTTTATCATGCT 1134  
 295 .....AsnSerHisIleGlySerThrAla.....ValArgLeu 305  
 1135 GCAGGTGGGTCAACAGTTATCGACCCAGACTGAATAATGCGACAAATAT 1184  
 306 AlaGlyAsnGluLysAsp.....AlaAsnGlyGlnAsnVal 318  
 1185 TTCCTTTATTACAAAGGAAAGGTGAATTGATACTTACCAGCAACATCA 1234  
 318 lThrPheGluAsp.....AsnGlyThrLeuValLeuAspGlnAsnIle 333  
 1235 ACCAAGCGCGGGCGTGTATTTCAGGGTAATTTTACGGTCTCGCCT 1284  
 333 snGlnGlyAlaGlyLeuPhePheLysGlyAspTyrThrValLysGly 349  
 1285 AAAACAACAGAA...ACGTGCCAAGGCGCGGCTTCATATCAGTGATGG 1331  
 350 AlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspG 366  
 1332 CAGTACCGCTTACTTGGAAAGTAAACGGCGTGGCAACAGCAGCGCTGCCA 1381  
 366 yLysLysValValThrGlnValLysAsnProAsnGlyAspArgLeuAla 383  
 1382 AAATCGGCAAAAGCAGCGTGTGTTCAAGCCAAGGGGAAACCAAGGC 1431  
 383 ysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGly 399  
 1432 TCGGTACCGTGGCGGACGGTAAAGTCATCTAGATCAGCAGCGCGACGA 1481  
 400 GlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAsp 416  
 1482 TCAAGGCAAAAACAAAGCTTTAGTGAATCGGCTTGGTCAGCGCGCAGG 1531  
 416 rAsnGlnLysValGlnAlaPheSerGlnValGlyIleValSerGlyArg 433  
 1532 GGACGTCGCACTGAATCCGATATCAGTTCAACCCCGACAACTCTAT 1581  
 433 rgThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyr 449  
 1582 TTCGGCTTCGGCGGCGAGTTTGATTGAAGTCAGCGGCAATTCGCTTCGT 1631  
 450 PheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPh 466  
 1632 CCACGCGATTCAAAATACCGATGAAGGGCGCGATGTTGTCACCAACAATC 1681  
 466 eGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAsn 483  
 1682 AAGACAAGAATCCCGTTACCATTACAGGCAATAAAGATATTACTACA 1731  
 483 hrClyHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAla 499  
 1732 ACGGCAAT 1740  
 500 ProGlnAsn 502

see documentation block:

A: Variety: SM1166  
C: Date: 23-Jul-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
C: Accession: S61330  
R: Lomholt, H.; Poulsen, K.; Mogens, K.





346 AACGGCGGCTATACAAATGTGTGATTTT.....GGTGC 377  
 51 AlahisGlnTyMetAsn...AspPheTyPheGlyHisTyAsnGlyHi 66  
 378 GGAGGAAGAACAATCCCGATCAGCACGCCGCTTTCTTACCRAAATGTCAAAA 427  
 :::: :::: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:  
 66 sArgAspValSerAspAspGluAsnLys.....TySerValValThrG 81  
 428 GAAATAATATAAACAGGAGGACTTAACGGCCAT.....CCTTATGCGGC 471  
 :|||:||||: :::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 81 InAsnAsnValAsnProAsnGluAsnTrpHisValAspLysArgLeuAsp 97  
 472 GATTATCATATCCCGCTTGTCACAAATTTGTCACAGATGCAGAACCTGT 521  
 |||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:  
 98 AspTyAsnMetProArgLeuAsnLysPheValThrGluValAlaProTh 114  
 522 TGAGATGACCACTGATTATGGATGGGTGGAAAA..TACGCTGATTAAATA 568  
 |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 114 rThrProThrLeuAlaGlyAspAspLeuGluThrTyLysAspLysGluL 131  
 569 AATACCTGATCGTGTCCAATTCGAGCAGGACAGCAA.....TATTGG 612  
 |||: |||: :::: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||: |||:||||:  
 131 ystyrLeuSerPheValArgValGlyAlaGlyArgGlnLeuValTyrglu 147  
 613 CGGCTCTGATCAA....GACGAACCCAAATAACCGGCAAGTTCATATCA 656  
 148 LysGlySerHisHisValGluAspLysGluHisGlyGluAspLeuLysAs 164  
 657 TATTGCAACGGCATATCTTGGCTCGTGGTGGCAATACCTTTGCA.... 702  
 :::::||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 164 pleuSerAlaIleTyArgTyrAlaIleGlyThrProTyLysGlyI 181  
 703 .....CAAATGGATCAGGTGGTGGCAGCAGTCACACTTAGTAGTAC 741  
 :::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 181 leasnIleaspProserGlnSerLysGlyLeulleGlyPheGlyAsp 197  
 742 GA AAAAATAAATACAT..... 756  
 :::::: |||  
 198 SerArgGluAspHisValIleAsnSerLysThrLeuLeuSerGlnAspPr 214  
 757 ...AGCCCATATGTTTTTACCACACAGGAGGCTCATTTGGCAGCAGTG 802  
 :::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 214 oueuthrasntyrglyvalLeu.....GlyaspSerg 225  
 803 GCTCACC AATGTTTATCTATGATGCCCAAGCAAAAGTGGTTAATTAAT 852  
 |||:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 225 lyserProLeuPhealpheasplysGlnGlnAsnLysTrpValPheile 241  
 853 GGGGTATTGCAACAGGCAACCCCTATATAGAAAAACCAATGGCTTCCA 902  
 |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 242 GlyProTyThrTyTrpalaGlytyrglyLysSerTrpGlnGluTr 258  
 903 GCTAGTTCGTAAGATGGTCTATGATGAATCTTTGCTGGAGTACCC 952  
 :::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 258 pAsnIleTyryLysSerGlnPheThrLysAspValLeuAsnLysAspSera 275  
 953 ATTCAGTATTCAGACCCACATCAAATGGGAATACTTTTTTAACGAC 1002  
 :|||:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 275 laGlyLeuLeu.....LysGlyAsnThrGlnTyAsnTrpThrSer 288  
 1003 AATAATAATGCGCAGGAAAAATCGATGCCCAACATAAACACTATTCTCT 1052  
 :|||:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 289 AsnGlyAsnThrSer..... 293  
 1053 ACCTTATAGATTAATAAACACGACCGTTCAATTTGTTTAATGTTTCTTTAT 1102  
 :::: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 294 .....MetIleSerAsnGlySerGluLeuLeuGluValAsnLeup 307  
 1103 CCGAGACGACGAAGAGACCTGTTTATCATGCTCGAGGTGGGTCAACAGT 1152  
 :|||:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:  
 307 heaspAsnSerLys.....HisThr 313  
 1153 TATGACCCACAGACTGAATAATGGAGAAAAATTTCCCTTTTATTCACAAAG 1202

seq\_documentation\_block:  
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (strain N; Alternate names: Iga1 protease  
C:Species: Haemophilus influenzae  
A:Variety: HK635  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
C:Accession: S61332  
R:Lomholt, H.; Poulsen, K.; Mogens, K.  
Mol. Microbiol. 15, 495-506, 1995  
A:Title: Comparative characterization of the Iga gene encoding Iga1 protease in Neisseria  
A:Reference number: S61314; MUID:95302961

```

alignment_scores:
  Quality: 718.50
  Ratio: 2.267
  Percent Similarity: 63.527
  Length: 499
  Caps: 17
  Percent Identity: 36.072

alignment_block:
  US-09-303-518D-653 x S61332 ..
  Align seg 1/1 to: S61332 from: 1 to: 462

```

```

202  GCGAAAGATATTAGGTTTACACAAAAAGGGGAGTGTGTGGCGCAATC 251
    |||  :::::|||||  :::::|||||  :::::|||||  :::::|||||  :::::
1  AlaThrAsnValGluValArgAspLysAsnAsnHisSerLeuGlyAsnAl 17

252  GATGACGGAAGCC...CGATGATTGATTTTCTGTGTTACGCGTAAC. 297
    ::::  :::::  |||||  |||||  |||||  |||||  |||||  |||||
17  aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34

298  ..GGCGTGGCGGCATTCGCGGGCGATCAATATATTGTGAGCTGGCACAT 345
    :::::  :::::  |||||  |||||  |||||  |||||  |||||  |||||
34  ysArgIleGlyThrLeuValAspProGlnTrpIleValSerValLysHis 50

```

252 GATGACGAAAGCC...CCGATGATTGATTTTCTGTGGTATCGCGTAAC. 297  
17 aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34  
298 ..GGCGTGGCGGCAATGGCGGGCGATCAATATATTTGTAGCGTGGCAGAT 345  
34 ysArgIleGlyThrLeuValAspProGlnTyrIleValSerValHis 50  
346 AAC.....GGCGGCTATAACAATGTTGA 368  
51 AlaHisLysHisIleAsnAspPheTyrPheGlyHisTyrAsn..... 64  
369 TTTTGGTGGCGGAGGAAAGCAATCCGATCAGCACCGCTTTTCTTACCAA 418  
65 ....GlyHisArgAspValSerAspGluAsnLys.....TyrSerV 78  
419 TTGTGAAAAGAAATTAATATAAGACGAGGACTAACGGCCAT.....CCT 462  
78 alValThrGlnAsnValLysProAsnGluAspTyrPheHisValAspLys 94  
463 TATGGCGGCGATTTATCATATCCGCGTTTCACAAATTTGTACAGATGC 512  
95 ArgLeuAspAspTyrAsnMetProArgLeuAsnLysPheValThrGluVa 111  
513 AGAACCTGTTGAGATGACCATTTATATGGATGGGTGGA....TACGCTG 559  
111 lAlaProThrThrProThrLeuAlaGlyAspAspLeuGluThrTyrLysA 128  
560 ATTTAAATAAATACCTGATCGTTCGATCGGAGGAGGAGGAGGAGGAGGAG 609  
128 sPlysGluLysTyrProSerPheValArgValGlyAlaGlyThrGlnPhe 144  
610 TGGCGGCTGAT.....GAAGACGAAACCAATACCGCGA 644  
145 ValTyrGluLysGlySerTyrTyrValGluLysThrThrArgAsnAsnAs 161  
645 AAGTTCATATCATATTCACCGCATATTTGGTCTCGTGGTGGCAATA 694  
161 pIleLysPhe...LeuAspGluAlaTyrArgTyrAlaIleGlyGlyThrP 177  
695 CTTTGA.....CAAATGGATCAGGTGGTGGCAGCAGTC 729  
177 roTyrGluGlyIleAsnLysAspProSerGlnSerLysLysGlyLeuL 193  
730 AACTTAGTAGCGAAGAAATTAACAT..... 756  
194 GlyPheGlyAspSerArgGluAsnHisValIleAspAlaLysThrLeuL 210  
757 .....AGCCCATATGGTTTTCACCAACAGGAGGCTCAT 790  
210 uSerGlnAspProLeuThrAsnTyrGlyValLeu..... 221  
791 TTGGCGACAGTGGCTCACCATGTTTATCTATGATGCCCAAGCAAGCAAG 840  
222 ..GlyAspSerGlySerProLeuPheAlaPheAspLysGlnGlnAsnLys 237  
841 TGGTTAATTAATGGGTGTTTGCACAAAGGCAACCCCTAT..... 879  
238 TrpValPheIleGly.....ProTyrThrTyrTyrAl 248  
880 ....ATAGAAAGCAAT.....GGCTTCAGCTAGTTCGTAAGATT 919  
248 aGlyTyrGlyLysLysSerTrpGlnGluTrpAsnIleTyrLysLysAsp. 264  
920 GCTTCTATGATCAATCTTGTCTGGAGATACCCATTTCAGTATTCTACGAA 969  
265 ..PheAlaAspAsnLysLysLysArgAspAsnAlaGluAlaVal..... 278  
970 CCACATCAAAATGGGAATACTTTTTTACGACAATAAATATGCGCAGG 1019  
279 ProPheSerThrSerGluTyrHisTrpThrAsnThrThrAsnHisGlnSe 295

314 AsnArgGluLysAlaAsnTyrGlyLysSerValThrPhe.....GlnG 328  
1203 AAAAGGTGAATTGATACATCAACGACCAATCAACAGCGCGCGGTT 1252  
328 yAsnGlyThrLeuThrLeuLysAsnSerIleAsnGlnGlyAlaGlyGlyL 345  
1253 TGTATTTGAGGTAAATTTACGGTC....TCGCCTAAACAAACGAAACG 1299  
345 euPhePheGluGlyAsnTyrThrValGluGlySerSerAspAsnIleVal 361  
1300 TGGCAAGCGCGCGGCTCATATAGTATGATGATGATGATGATGATGATG 1349  
362 TrpAsnGlyAlaGlyIleSerValAlaGluGlyLysThrValThrTrpL 378  
1350 AGTAAACGCGGTGGCAACAGCGGCTGTCCAAATCGGCAAGGACGCG 1399  
378 sValHisAsnProGlnSerAspArgLeuAlaLysIleGlyLysGlyThrL 395  
1400 TCGTGGTCAACGCAAGGGAACCAAGGCTCGTCCGTCAGGTCGGCGAC 1449  
395 euIleValGluGlyLysGlyGluAsnLysGlySerLeuLysValGlyAsp 411  
1450 GGTAAAGTTCATCTAGATCAGCAGCGGACGATCAAGGCAAAAAACAAGC 1499  
412 GlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLysValLysAl 428  
1500 CTTTGTAGTAAATCGGTCGTCAGCGGACGAGGAGGAGGAGGAGGAGGAG 1549  
428 aPheSerGlnValGlyIleValSerGlyArgSerThrValValLeuAsnA 445  
1550 CCGATAATCAGTTCACCCCGCAACTCTATTCGGCTTCGCGGC 1596  
445 spAspLysGlnValAspProAsnSerIleTyrPheGlyPheArgGly 460  
seq\_name: pir2.S61333  
seq\_documentation\_block:  
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Haemophilus influenzae (strain N)  
N:Alternate names: Iga1 protease  
C:Species: Haemophilus influenzae  
A:Variety: HK284  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
C:Accession: S61333  
R:Lomholt, H.; Poulsen, K.; Mogens, K.  
Mol. Microbiol. 15, 495-506, 1995  
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria meningitidis serotype 4  
A:Reference number: S61314; MUID:95302961  
A:Accession: S61333  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-471 <LOW>  
A:Cross-references: EMBL:X82487; NID:g773205; PIDN:CAA57870.1; PID:g773206  
A:Experimental source: strain HK284  
C:Genetics:  
A:Gene: iga  
C:Superfamily: Iga-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase  
alignment\_scores:  
Quality: 695.50 Length: 514  
Ratio: 2.153 Gaps: 22  
Percent Similarity: 62.840 Percent Identity: 34.825  
alignment\_block:  
US-09-303-518D-653 x S61333 ..  
Align seg 1/1 to: S61333 from: 1 to: 471  
202 CGCAAGATATTAGGTTTACACAAAGAGGGAGTGTGTCGGCAATC 251  
1 AlaThrAsnValGluValArgAspLysAsnAsnGlnSerLeuGlySerAl 17

```

1020 AAAATCGATGCCAAACATAACACTATTCTCTACCTTATAGATTAATAA 1069
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
295 rGlulle.....LysAsnThrAspHisThrIleThrValThrLeuProS 310
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1070 CACGAACCGTTCAATTTGTTTAAATGTTCTTATCCGAGACGACGAGAA 1119
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
310 erAspProAsnArgLeuValAsnPhelGlnLysGlu..... 322
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1120 CCTGTTTATCATGTCGAGTGGGGTCAACAGTTATCGACCCAGACTGAA 1169
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
323 .....HisLeuG1 325

1170 TAATGAGAAAAATATTCCTTTATTGAC.....AAAGGAAAGGTC 1210
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
325 nThrGlyGlnAsnValThrPheAspSerThrAsnAsnGlyLysGlyT 342
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1211 AATTGATACCTACGACGACATCAACCAAGCGCGCGGCTTGTATTTT 1260
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
342 hrLeuIleLeuAspAspHisIleAsnGlnGlyAlaGlyLeuPhePhe 358
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1261 GAGGCTAATTTACGGTCTCCCTTAACACAGAA...ACGTGGCAAGG 1307
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
359 LysGlyAsnTyrGluValLysGlyLysThrAspAlaThrTrpValG1 375
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1308 CGCGGGCGTTCATATCAGTCGACGAGTACCGTCTTCTGGAAGTAAACG 1357
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
375 YGlyGlyIleAspValAlaGluGlyLysGluValValTrpLysValHis 392
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1358 GCGTGGCAACGACCGCTGTCCAAATCGCAAGGACGCTGTGCTGTT 1407
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
392 snProGluLysAspHisLeuAlaLysIleGlyLysGlyThrLeuIleVal 408
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1408 CAAGCCAAAGGGGAAACCAACGAGTCTGTCAGCGGCGGCGGCTAAAGT 1457
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
409 GluGlyLysGlyAsnAsnLysGlySerLeuLysValGlyAspGlyThrVa 425
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1458 CATCTTAGATCAGCAGCGCGGACGATCAAGGCAAAACAAAGCTTACTG 1507
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
425 lValLeuLysGlnGlnThrAsnGlySerGly...GluHisAlaPheAlas 441
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1508 AAATCGGCTGTGTCAGCGGCGGAGCGGTGCAACTGAATCCGATAT 1557
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
441 erValGlyIleValSerGlyArgSerThrValValLeuAsnAspLys 457
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1558 CAGTTCACCCGACAACTCTATTCGCTTTCGCGCGGCA 1599
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
458 GlnValAspProAsnSerIleThrPheGlyPheArgGlyGly 471
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||

```

seq\_name: pir2:T00317

```

seq_documentation_block:
probable serine proteinase espP, extracellular - Escherichia coli plasmid p0157
N:Alternate names: putative exoprotein-precursor
C:Species: Escherichia coli
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Dec-2000
C:Accession: T00317; T42120
R:Maikino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota, S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli O157:H7, substrain RIMD 0509952
A:Reference number: 22068; MUID:98391744
A:Accession: T42120
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli O157:H7, substrain RIMD 0509952
A:Reference number: 22068; MUID:98391744
A:Accession: T42120
A>Status: preliminary; translated from GB/EMBL/DBJ

```

```

A:Molecule type: DNA
A:Residues: 1-1300 <BUR>
A:Cross-references: EMBL:AF074613; PIDN:AACT0088.1
A:Experimental source: strain EDL933; serotype O157:H7
C:Genetics:
A:Gene: espP
A:Genome: plasmid p0157
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: serine proteinase

```

```

alignment_scores:
Quality: 629.50 Length: 1505
Ratio: 0.844 Gaps: 64
Percent Similarity: 49.568 Percent Identity: 21.794

```

alignment\_block:

US-09-303-518d-653 x T00317 ..

Align seg 1/1 to: T00317 from: 1 to: 1300

```

82 TTAGCCATATGCCTGCTCGTTCGGCATTCGCCCAAGCCCGCGGGGACA 131
      ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
40 LeuAlaLeuCysPhe...LeuGlyLeuLeuGlnSerTyrSerPheAl 55
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
132 CACTTATTTCGGCATCAACTACCAATAC.....TATCGCGACTTGG 172
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
55 aSerClnMetAspIleSerAsnPheTyrIleArgAspTyrMetAspPhe 72
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
173 CCGAAATAAGGCAAGTTTTCAGTCGCGGGCGAAAGATATTGAGTTTAC 222
      ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
72 laGlnAsnLysGlyIlePheGlnAlaGlyAlaThrAsnIleGluIleVal 88
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
223 AACAAAAGGGGAGTTGGTCGGCAATCGATGACGAAAGCCCGGATGAT 272
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
89 LysLysAspGlySerThrLeu.....LysLeuProGluValProPhePr 103
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
273 TGATTTTCTGTGATCGGTAAACGGCTGCGGGCGCATTCGCGGGGATC 322
      ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
103 oAspPheSerProValAlaAsnLysGlySerThrThrSerIleGlyGly 120
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
323 AATATATTGTGCGGTGGCACATAACGGCGGCTATAACAATGTTGATTTT 372
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
120 laTyrSerIleThrAlaThrHisAsn..... 128
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
373 GGTGCGGAGGGAAGCAATCCGATCCGATCAGCACCGCTTTCTTACCAATTGT 422
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
129 .....ThrLysAsnHisHisSerValAlaThrGlnAsnTr 140
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
423 GAAAAGAAATAATATAAGCA.....GGGACTAACGGCCCTCTTATG 466
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
140 pGlyAsnSerThrTyrLysGlnThrAspTrpAsnThrSerHisPro.... 155
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
467 GCGGCGATTCATATATCCCGCTTTCGACAAATTTCTCAGATGACGAGAA 516
      ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
156 .....AspPheAlaValSerArgLeuAspLysPheValValGluThrArg 170
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
517 CCGTTCAGATGACCACTATATATGATGGTGGAAATACGCTGATTATAA 566
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
171 GlyAlaThrGluGlyAlaAspIleSerLeuSerLysGlnGlnAlaLeuG1 187
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
567 TAAATACCTGATCGT.....GTTCTGAA 589
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
187 uArgTyrGlyValAsnTyrLysGlyGluLysLysLeuIleAlaPheArgA 204
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
590 TCGGACGAGGACACATAATTGGCGTCTGATGAGACGACCAACCAATAC 639
      ||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
204 laGlySerGlyValValSerValLysLysAsnGlyArgIleThrProPhe 220
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
640 CGGGAAGTTCATAT.....CATAT 659
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
221 AsnGluValSerTyrLysProGluMetLeuAsnGlySerPheValHisI1 237
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||

```



660 TCACAGCGCATATCTTGGCTGCTC...GGTGGCAATACCTTTGCACAAA 706  
 1 ...  
 237 eAspAspTrpSerGlyTrpLeuIleLeuThrAsnAsnGlnPheAspGluP 254  
 1 ...  
 707 ATGGATCAGGTGGTGGCAGCAGTCAACTTAGGTAGCGAAGAAAATTAACAT 756  
 1 ...  
 254 heAsn...AsnIleAlaSerGln... 260  
 1 ...  
 757 AGCCCATATGTTTATACCAAGAGGCTCATTTGGCGACAGTGGCTC 806  
 1 ...  
 261 ...GlyAspSerGlyse 265  
 1 ...  
 807 ACAATGTTTATCTATGATGCCCAAGCAAGTGGTTAATTAATGGG 856  
 1 ...  
 265 rAlaLeuPheValTyrAspAsnGlnLysLysTrpValValAlaGlyT 282  
 1 ...  
 857 TATTGCAACAGGCAACCCCTATATAGGAAAAGCAATGGCTTCCAGCTA 906  
 1 ...  
 282 hrVal... 283  
 1 ...  
 907 GTTCGTAAGATTGGTTCTATGATGAATCTTTGGGAGATACCCATTC 956  
 1 ...  
 284 ...TriplyIleTyrAsnTyrAlaAsnGlyLysAsnHisAl 296  
 1 ...  
 957 ACTATTCTACGACCAATCAA...AATGGCAAT 988  
 1 ...  
 296 aAlaTyrSerLysTrpAsnGlnThrIleAspAsnLeuLysAsnLysT 313  
 1 ...  
 989 ACTTTTAAACGCAATAATATGGCGCAGGAAATTCGATCCCAACAT 1038  
 1 ...  
 313 yrSerTyrAsnValAspMetSerGlyAla... 322  
 1 ...  
 1039 AAACACTATTCTTACCTATAGATTAACACAGCAACGCTCAATGTT 1088  
 1 ...  
 323 ...GlnValAlaThrIleGlu... 328  
 1 ...  
 1089 TAATGTTTCTTTATCCGAGACCAAGAACCTGTTTATCATGCTGCAG 1138  
 1 ...  
 329 .AsnGlyLysLeuThrGlyThrGlySerAsp... 338  
 1 ...  
 1139 GTGGGTCAACAGTATTCGACCCAGACTGAATATAGGAAATATTTCC 1188  
 1 ...  
 339 ...ThrThrAspIleLysAsnLysAsp 346  
 1 ...  
 1189 TTTATTGACAAAGAAAGCTGAATTCATCTTACCAGCAACATCAACCA 1238  
 1 ...  
 347 LeuIlePheThrGlyGlyGlyAspIleLeuLeuLysSerPheAspAs 363  
 1 ...  
 1239 AGCGCGGGGGTGTGTTATTTGAGGGTAATTTACGGTCTCCCTAATAA 1288  
 1 ...  
 363 nGlyAlaGlyGlyLeuValPheAsnAspLysLysThrTyrArgValAsnG 380  
 1 ...  
 1289 ACAACGAA...ACGTGCAAGGCGGGCGGTTCATATCATGTATGTCAGT 1335  
 1 ...  
 380 lyAspAspPheThrPheLysGlyAlaGlyValAspThrArgAsnGlySer 396  
 1 ...  
 1336 ACCGTACTTGTGAAGTAAGCGGTGGCAACAGCCGCTGTCCAAAT 1385  
 1 ...  
 397 ThrValGluTrpAsnIleArgTyrAspAsnLysAspAsnLeuHisLysII 413  
 1 ...  
 1386 CGGCAAGGCGACGTGTGTTCAAGCAAGGGAAGGAAACCAAGGCTCGG 1435  
 1 ...  
 413 eGlyAspGlyThrLeuAspValArg...LysThrGlnAsnThr...AsnL 428  
 1 ...  
 1436 TCAGCGTGGGCGAGGTAAAGTCACTTAGATACAGCAGCGGACGATCAA 1485  
 1 ...  
 428 euLysThrGlyGluGlyLeuValIleLeu... 437  
 1 ...  
 1486 GGCAAAACCAAGCTTTAGTGAATCGGTTGTCAGCGCAGGGGAC 1535  
 1 ...  
 438 GlyAlaGluLysThrPheAsnAsnIleTyrIleThrSerGlyAspGlyTh 454  
 1 ...  
 1536 GGTGCACTGAATGCCGATATCATGTTCAACCCCGACAAA...C 1576

454 rValArgLeuAsnAlaGluAsnAlaLeuSerGlyGlyGluTyrAsnGlyI 471  
 1 ...  
 1577 TCTATTTCGGCTTTTCGGCGGAGCTTTGGATTGTAACGGCGCATTCGCTT 1626  
 1 ...  
 471 lePhePheAlaLysAsnGlyGlyThrLeuAspLeuAsnGlyTyrAsnGln 487  
 1 ...  
 1627 TCGTTCCACCGCATTCAAATACCGATGAAGGGCGATGATTGTCAACCA 1676  
 1 ...  
 488 SerPheAsnLysIleAlaIleThrAspSerGlyAlaValIleThrAsn... 503  
 1 ...  
 1677 CAATCAAGACAAAGAATCCCGCTTACCATTACAGCAATAAAGATATTA 1726  
 1 ...  
 504 .ThrSerThrLysLysSerIleLeuSerLeuAsnAsnThrAlaAspTyrI 520  
 1 ...  
 1727 CTACACCGGCAAT...AACACAACTTGGATAGCAAAAGAAATTTGCC 1773  
 1 ...  
 520 leTyrHisGlyAsnIleAsnGlyAsnLeuAspValLeuGln... 533  
 1 ...  
 1774 TACAACGGTTGGTTGGCGAGAAAGATGCAACCAAAACGAACGGCGGCT 1823  
 1 ...  
 533 ... 533  
 1 ...  
 1824 CAATCTGAATTTACCAACCGGAAGAGCGGATCGCACTTTACTGTTTCCG 1873  
 1 ...  
 534 .....HisHisGluThrLysLysGluAsnArgArgLeuIleLeuAspG 548  
 1 ...  
 1874 CGGNACAAATTTAAACCGCAATATCAGCAACAAACGCGCAACTGTTT 1923  
 1 ...  
 548 lyGlyValAspThrThrAsnAspIleSerLeuArgAsnThrGlnLeuSer 564  
 1 ...  
 1924 TTCACGGCAGACCGCACACCGCCCTACCAATCATTTAGGAAGC... 1968  
 1 ...  
 565 MetGlnGlyHisAlaThrGluHisAlaIleTyrArgAspGlyAlaPheSe 581  
 1 ...  
 1969 .....GGTGTGCARAAATGG 1984  
 1 ...  
 581 rCysSerLeuProAlaProMetArgPheLeuCysGlySerAspTyrValA 598  
 1 ...  
 1985 AAGGTATC.....CCACAAGGAGAAATCGTGTGGGCAAC... 2019  
 1 ...  
 598 laGlyMetGlnAsnThrGluAlaAspAlaValLysGlnAsnGlyAsnAla 614  
 1 ...  
 2020 .....GATTGGATCGA 2030  
 1 ...  
 615 TyrLysThrAsnAsnAlaValSerAspLeuSerGlnProAspTrpGluTh 631  
 1 ...  
 2031 CGGCACATTTAAAGCGGAAACTTCCATATTCAGGCGGACAAAGCGGTGG 2080  
 1 ...  
 631 rGlyThrPheArgPheGlyThrLeuHisLeuGluAsnSerAspPheSerV 648  
 1 ...  
 2081 TTTCCCGCAATGTGCCAAAGTGAAGCGGATTTGGCATTTAAGCAATCAC 2130  
 1 ...  
 648 alGlyArgAsn...AlaAsnValIleGlyAspIleGlnAlaSerLysSer 663  
 1 ...  
 2131 GCCCAAGCAGTT..... 2142  
 1 ...  
 664 AsnIleThrIleGlyAspThrThrAlaTyrIleAspLeuHisAlaGlyLy 680  
 1 ...  
 2143 .....TTCGGTGTGCGACCGCAT... 2160  
 1 ...  
 680 sAsnIleThrGlyAspGlyPheGlyPheArgGlnAsnIleValArgGlyA 697  
 1 ...  
 2161 .....CAAAGCCACACAATCTGTACAGTTCGGACTGGACGGGTCTGACA 2205  
 1 ...  
 697 snSerGlnGlyGluThrLeuPheThr.....GlyGlyIleThr 709  
 1 ...  
 2206 AGTTGTACCGAAGAAAACCATATTCCGACGATAAAGTGAATTCCTTCATGAG 2255  
 1 ...  
 710 AlaGluAspSerThrIleValIleLysAspLysAlaLysAlaLeuPheSe 726  
 1 ...  
 2256 C.....AAGACCGACATCAGAGGCAATGTCA 2281  
 1 ...







[illegible]

1063 LeuGlyAsnGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSerAr 1079  
2835 .....  
1079 gGlySerValTrpLeuAsnSerAsnSerCysAlaGlyThrSerAsnC 1096  
2835 .....  
1096 ySgluTyArgValAsnSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1112  
2836 .....  
1113 SerAlaGlnThrAlaAlaProAlaThrAsnGlyIleTyArgAsnThrLe 1129  
2868 CACGGTAAACGGCAAAATGACGTCAGGCAACATTCGGCTTTATGTCGG 2917  
1129 uThrAsn...GluLeuSerGlySerGlyAsnPheTyLeuHisThrA 1145  
2918 AACTCTTCGGCTACCGCAGCGCAAAATGAGCTGGCGGAAAGTTCGGAA 2967  
1145 snValAlaGlySerArgGlyAspGlnValValAsnAsnAlaThr 1161  
2968 GGCACTTACACCTTGCTGTCAACAATACCGCAACGCAACCCCTAAGTCT 3017  
1162 GlyAsnPheLysIlePheValGlnAspThrGlyValSerProGlnSerAs 1178  
3018 CGACCAATTACCGGTAGTGGAA.....GAAAAGACACACACCCCTGT 3061  
1178 pAspAlaMetThrLeuValLysThrGlyGlyGlyAsp..... 1190  
3062 CCGAAAATCTTAATTCACCTCTCAAAACGAA.....CACGTGATGCC 3105  
1191 .....AlaSerPheThrLeuGlyAsnThrGlyGlyPheValAspLeu 1204  
3106 GGGCGATGCGGTATACGCTTATCCCAAGAGCGGAGTTCGCGCTGCA 3155  
1205 GlyThrTyrgluTyf...ValLeuLysSerAspGlyAsnSer..... 1217  
3156 TAATCCGGTCAAGAACAGAGCTTTCGACAACTCGCAAGCGGGAG 3205  
1217 ..... 1217  
3206 AAACAGAGCGCGCTTGACGGCAAAACAGGCACAACTTCCGCCAAACAA 3255  
1217 ..... 1217  
3256 CAGCGCGAAAGACAAACCGCGCAAGCCTTGACGCGCTGATTCGGCGCG 3305  
1218 .....AsnTrpAsnLeuThrAsnAspValLysProA 1228  
3306 GCGCAATGCCACGAAAGAGCAGAAAGTTCGCCGAACCGCGCGCGAGG 3355  
1228 snProAspPro.....IleProAsn..... 1234  
3356 CAGCGCGGGAATAATCGCGCATTTATGACGGCGGAGGAAAGAAAAACGG 3405  
1234 ..... 1234  
3406 GTCCAGCGCGGATAAGACACCCCTTGGGAAACAGCGCGAAGCGGAAC 3455  
1235 .....ProLysp 1237  
3456 CCGCGCGGCTACCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGGATT 3505  
1237 roAspProLysProAspProLysProAspPro...AsnProLysProAsp. 1252  
3506 TCGCCGAACCGGACCCCAACCGCAACCCCAACCGCAGCGAGCTGATC 3555  
1253 ..ProThrProAspProThrProThrProValProGluLysArgIleThr 1268  
3556 AGCCCGTTATGCAATAGCGGTTTNGTGAATTTCCGCGCAGCTCAACAG 3605  
.....

1269 Pro.....SerThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLe 1283  
3606 CGTTTTC.....GCCGTACAGGACGAATTTGGACCGCGTGT 3640  
1283 uValPheAspAlaGluLeuAsnSerIleArgGluArgLeuAsnIleMetL 1300  
3641 TTGCCGAAGACCGCGCAACGCGTTTGGACAAGCGGCATCCGGACACCC 3690  
1300 yAlaSerProHisAsnAsnValTrp...GlyAlaThrTyArgAsnThr 1315  
3691 AAACACTACCGTTCGCAAGATTTCGCGCGCTACCGCCCAACAAACCGACCT 3740  
1316 ArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGluGlnThr..... 1330  
3741 GCGCCAAATCGGTATGCGAAGAAACCTCGGCAGCGGCGC.....G 3781  
1331 ....LeuThrGlyMetThrValGlyIleAspSerArgAsnAspIleProG 1346  
3782 TCGGCATCTCG.....TTTTCGCAACACCGGACCC 3810  
1346 luGlyIleThrThrLeuGlyAlaPheMetGlyTyArgHisSerHisIle 1362  
3811 GGAACACCTTCGACGAGCGCATCGCAACTCGGCACGGCTTGCCACCG 3860  
1363 Gly.....PheAspArgGly.....GlyHisG1 1370  
3861 TGCCCTTTTCGGGCAATACGGCATCGGC..... 3888  
1370 ySerVal...GlySerTyArgSerLeuGlyGlyTyArgAlaSerTrpGluHisG 1386  
3889 .....AGTTTCGACATC 3900  
1386 luSerGlyPheTyArgLeuAspGlyValValLysLeuAsnArgPheLysSer 1402  
3901 GGCAATCAGC.....GCGGGCGCGGTTTGTAGTACGCGCAGCCCTT.. 3939  
1403 AsnValAlaGlyLysMetSerSerGlyGlyAlaAlaAsnGlySerTyR1 1419  
3940 .TCACAGCGCATCAGAGCAAAATCCGCGCGCGCTGCTGCATTACGGCA 3988  
1419 sSerAsnGlyLeuGlyHis.....I 1427  
3989 TTCAGGCAAGATACCGCGAGGTTTCGGCGGATTCGCATCGAACCCGAC 4038  
1427 leGluThrGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTy 1443  
4039 ATCGCGCAACCGCGCTATTTCGTCCAAAACCGCATTTACCGATACGAAA 4088  
1444 AlaSerLeuThrGlyPheThrAlaAspAsnProGluTyArgHisLeuSerAs 1460  
4089 C.....GTCAATATCGCCACCGCGCGCTTGCATTCAACCGCT 4126  
1460 nGlyMetLysSerLysSerValAspThrArgSerIle.....T 1473  
4127 ACCCG...GCGGCGATTACGCGAGATTTATTCATTCAACCGCGCAACAC 4173  
1473 yrArgGluLeuGlyAlaThrLeuSerTyArgAsnMetArgLeuGlyAsnGly 1489  
4174 ATTTCCATCATCGCTTATTAGCGCTGTCTCTATACCGATGCCCTTCCGG 4223  
1490 MetGluValGluProTrpLeuLysAlaAlaValArgLysGluPheValAs 1506  
4224 CAAAGTCCGAACCGCGCTCAATACCGCGCTTATTTGGCGCAGGATTC... 4269  
1506 pAspAsnArgValLysValAsnSerAspGlyAsnPheValAsnTyArgLeuS 1523  
4270 ..GGCAAAACCCGAGTGGGAATGGGGCTAAACCCGGAATCAAGGT 4317  
1523 erGlyArgArgGlyIleTyArgGlnAlaGlyIleLysAlaSerPheSerSer 1539  
4318 TTCACGCTGTCCCTCCACGCTGCCCGCGCAAGGGCGCGCAATTTGGAAC 4367  
1540 ...ThrLeuSerGlyHisLeuGlyValGly.....Ty 1549

```
4368 GGAGCACACGGCGGCATCAAA.....TTAGGCTACC 4399
:::|||||:::
1549 rSerHisSerAlaGlyValGluSerProTrpAsnAlaValAlaGlyValA 1566
4400 GCTGG 4404
1566 snrp 1567

seq_name: pir2:B90674

seq_documentation_block:
A:IdA-I adhesin-like protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
R:Accession: B90674
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90674
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1327 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA833785.1; PID:g13359819; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: ECs0362

alignment_scores:
Quality: 348.00 Length: 1373
Ratio: 0.509 Gaps: 72
Percent Similarity: 49.818 Percent Identity: 20.612

alignment_block:
US-09-303-518D-653 x B90674
Align seg 1/1 to: B90674 from: 1 to: 1327
1024 ATGATGCCAAACATAAACAATCTCTACCTATATAGATTAAACACG 1073
59 IleAspPheGlyLysGluAsnGlnSerIleAspThrArgIleThrValTh 75
1074 AACGTTCAATGTTTAATGTTCTTATCCGACAGCAGCAAGAACCTG 1123
75 rAspAsnAlaAsnLeuValIleAsnAlaThrAspThrSerArgProArgL 92
1124 TTTATCATGTCAGGTGGG.....GTCAACAGTTATCCAGCCAGA 1164
92 eutThrLeuAlaSerGlyGlyLeuAspIleThrGlyGlyLysValThr 108
1165 CTGAATTAATGAGAAATATTTCTTTATTGACAAAGGAAAGTGAT 1214
109 IleAsnGly.....ProLeuAsnPheLeuLeuLysGlyThrGlyPheLe 123
1215 GATACCTTACCAGCAACATCAACCAAGCGCGCGGTTTGTATTGAG. 1263
123 u.....AsnValSerAsnAlaGlySerGluLeuTyAlaAspA 136
1264 .....GTAATTTTACG 1275
136 spLeuTyGluSerAsnSerGlyMetArgHisAspArgGlyTyPheAsn 152
1276 GTCTCGCTAAACACAGCAACGTGGCAAGCGCGCGGCTTCATATCAG 1325
153 ValSer.....AsnGlyGlyLysIleHisVally 162
1326 TGATGGCAGTACCGTTACTTGTAAAGTAAACGGCGTGGCAACACGCC 1375
162 scGlyThrSerArgLeuThrTyLeuGlnGlyAsnValSerGlyGlyGly. 178
```

```
1376 TGTCCAAATCGCAAGGACGCTGCTTCAAGCCAAAGGGGAAAC 1425
|||||:::
179 ..SerGlnValAsnSerGluThrPhePheMetGlyValTyGlySerTy 194
1426 CAAGGCTCG.....GTCACGCTGGCGACGGTAAAGTCATC..... 1461
|||||:::
195 GlyGlyAsnGlnTyLeuSerValAsnAsnGlyGlyGluValAsnAla 211
1462 .....TTAGATCAGCAGCGCGGACGATCAAGGCA 1489
211 gLysGlnIleSerLeuGlyTyTyAspGlnValSerAsp.....T 225
1490 AAAACAAGCCTTTAGTGAATCGCTTGTGACGGCAGGGGACGGT 1539
||| ||||| ||| ::|||::: |||:::
225 hrThrLeuAlaValSerGluGlyLysIleSerAla...ProThrIle 240
1540 CAACCTGAATGCCGATATCATGTTCAACCCGACAACTCTATTTCG 1589
|||||:::
241 SerLeuSerThrAsnSerGlu.....LeuAlaLeuGlyAl 252
1590 TCGCGGGGACGCTTGGATTGAAACGGGCATTCGCTTTCCTCCACG 1639
::: ||| ::::: :::::
252 acGlnGluGlySerAlaAlaLysAlaAlaGlyIleIleAspAlaGly 269
1640 TTCAA.....AATACCGATGAAGGGCGGATGTTGTCAACAC 1677
269 leGluPheValTrpAlaLysThrSerGluLysLysIleThrLeuAsn 285
1678 AATCAAGACAAAGATCCACCGTTACC.....ATTACAGCAATAA 1718
::: |||||:::
286 Thr...AspLysAspAlaThrIleSerAlaAspIleValSerGly 301
1719 AGATATT.....ACTACA.....ACCGCA 1738
301 uGlyLeuGlyTyIleAsnAlaLeuAsnGlyThrThrTyLeuThrGly 318
1739 ATACACAACACTTGATAGCAAAAGAAATTCCTACACGCTTGTGTT 1788
::|||::: ::|||::: |||||:::
318 spAsnSerAlaPheSerGlyLysValLysIleGluGlnAsnGlyAla 334
1789 GCGAGAAAGATGCAACCAACGACGCGGCTCAATCTGAATTACCA 1838
||| |||||::: ||| ::|||:::
335 Gly.....IleThrGlnAsnIleGlyThrAlaGluIleAsnAsn 348
1839 ACCGAA.....GAAGCGGATCGCACTTACTGCTTCCCGCGAA 1879
348 gGlyLysLeuHisLeuLysAlaAspSerMetThrPheAlaAsnLys 365
1880 CAATTTAAACGGCAATATACGCAAAACGCAAACTGTTTTCACG 1929
::: |||||::: |||||:::
365 leSerGlyAsnGlyThrIleSerIleAspSerGlyThrValGluLe 381
1930 GGCAGACCGACACCGACGCTTACATCATTTAGGAGCGGTGG..... 1974
|||||::: |||||:::
382 GlyAsn.....AsnTyAlaPheSerGlyTyIleAs 392
1975 .....TCAAAATGGAAGGTATCCAC 1996
392 pValAlaSerGlyAlaValAlaValIleSerGluAspLysAsnIle 409
1997 AAGGAGAAATCGTGTGGAC.....AACGAT 2022
409 rGalaGluLeuAspValAspGlyLysLeuGlnIleAsnAlaAsnLys 425
2023 TGGATC...GACCGCACATTTAAACGGGAAACTTCCATATTTCAGG 2069
|||||::: |||||:::
426 TrpValPheAspAsnAspLeuGlu.....G 434
2070 ACAAGCGGTGTTTCCCGCAATGTTCCCAAGTGGNA..... 2106
434 yArgGlyIleValGluIleAsnMetGlyAsnHisGluPheSerPhe 451
2107 .....GCCGATTGG..... 2115
```

```

451 l uPheAlaTyrThrAspTrpPheGlnGlySerLeuAlaPheGlnAsnThr 467
|||||
2116 .....CATTTAAGCAATACAGCCCAAGCACTT.....TTCCGTGTGCGC 2153
|||||
468 ThrPheAsnLeuGluLysAsnAlaGluPheLeuGlnGlyGlyIleTh 484
|||||
2154 ACCGCATCAA.....AGCCACACAATCT 2176
|||||
484 rAlaGlyGlnGlySerLeuValThrValGlyLysGlyAlaHisSerIleS 501
|||||
2177 GTACAGCTGC.....GACTGGACGGTCTCAGCAAGT 2208
|||||
501 erThrLeuGlyPheSerGlyThrValAspPheGlyAlaLeuThAla 517
|||||
2209 .....TGTACCGAAAAACCATACCGACGATATAAGT..... 2241
|||||
518 GlyAlaGlnMetThrGluGlyThrValAsnValSerLysThrLeuAspLe 534
|||||
2242 .....ATTGCTTCATTGACAGACAGCCAGCATCAGAGGCAATG 2278
|||||
534 uArgGlyGluGlyValIleGlnValSerAspSerValValArgSerV 551
|||||
2279 TCAGCCTTGCGGATCAGCGCTCATTTAAATCTACA.....GGA 2316
|||||
551 alSerArgAspIleAspSerAlaLeuSerLeuThrGluValAspGly 567
|||||
2317 CTTCGCCACACTCAAC..... 2331
|||||
568 AsnSerThrIleLysLeuValAspAlaGlnGlyValGluValLeuGlyAs 584
|||||
2332 .....GGCAACTTAGTCA.....GGCGGACAGCCCATATATA 2365
|||||
584 pAlaGlyAsnLeuGlnLeuGlnAspLysAsnGlyGlnIleLeuSerSerS 601
|||||
2366 CGGTTACGCGCAACGCCCAACCAAGCACTCAGCTCGTGGGCAAT 2415
|||||
601 erAlaGlnArgAspIleGlnGlnAsnGlnLysAlaAlaValGly... 616
|||||
2416 GCCAAGCAACATTAATCAAGCCACATTAAAGCGCAACACATCGGCTTC 2465
|||||
617 .....ThrPyrAspTyrArgLeuThrSerGlyValAsnAsnAspG1 630
|||||
2466 GGCAATAGCTTCATTAACTTAGCAAC.....AACGCGGTACAAA 2506
|||||
630 yLeuTyrIleGlyThrGlyLeuThrGlnLeuAspLeuHisAlaThrAspS 647
|||||
2507 ACGGCACTGTAGCTTTCGACAAAGCTAAGGCAACGTAAGCCATTC 2556
|||||
647 erAspAlaLeuValLeuSerSerAsnGlyLys.....SerGluAsn 660
|||||
2557 GCACCTCAAGGCAATGCTCCCTAGCCGATAGGCAATTCATTTCAT 2606
|||||
661 AlaAlaAspLeuSerAlaLysIleThrGlySerGlyAspLeuAlaPheSe 677
|||||
2607 AAACAGCCGCTTACCGGAAAAATCAGCGCGCAAGGATACGGCATTAC 2656
|||||
677 rSerGlnLys.....GlyGlnThrValSerLeuS 687
|||||
2657 ACTTAAAGACAGCGAATGAGCGCTGCGCTCGGCGACGGAATAGGCAAT 2706
|||||
687 erAsnLysAspAsnAspTyrThrGlyValThrAspLeuArgSerGlyThr 703
|||||
2707 TTAAACCTTGACACGCCACCATTTACACTCAATTCGGCTATCGACAGA 2756
|||||
704 LeuLeuLeuAsnAsnAspAsnValLeuGlyAsnThr.....HisG1 717
|||||
2757 TCGCGCAGCGCGCAACAGCGGAGTGCAGATCGCGCGCGCCGCGCTT 2806
|||||
717 u..... 717
|||||
2807 CGCGCGCTTCCCTATATCCGTTAGCGCGCAACTTCGGCAGATCCCGT 2856
|||||

```

```

718 .....LeuArgLeuAlaAlaGluThrGluLeuAspMetAsn 729
|||||
2857 TTCAACACGCTGACGGTAACGGCAAAATTGAACGGTCAGGGAACATTCG 2906
|||||
730 GlyHisSerGlnThrVal...GlyThrLeuAsnGlySerAla..... 742
|||||
2907 CTTTATGTGCGAACTCTTCGGCTACCGCAGCGCAAAATTGAAGCTGGCG 2956
|||||
743 .....AspSerLeuLeuSerLeuAsnGlyGlySerLeuThrValThrA 757
|||||
2957 AAAGTTCGGAAGCACTTACACTTGGCTGTCAACAATACCGCAACGAA 3006
|||||
757 snGlyGlyThrSerThrGlySerLeu.....ThrGlySerGly 769
|||||
3007 CCGGTAAAGTCTCAGCAATTCAGCGTAGTGGAGGAAGAAAGACACACAC 3056
|||||
770 GluLeuAsnIleGlnGlyThrLeuAspIleAlaGlyAspAsnSerAs 786
|||||
3057 GGTGTCGCCGAAATCTTAATTTTC.....ACCGTCAAAACG 3091
|||||
786 nLeuThrAlaAsnValAsnIleAlaAsnSerAlaAsnValLeuValSerH 803
|||||
3092 AACAGTGCATCCCGCGCATGGCTTATCAGCTTATCCGCAAGACGCG 3141
|||||
803 isAlaGlnGlyLeuGlySerAlaAsn.....ValGluAsnAsnGly 816
|||||
3142 GAGTTCGCCGCTGCATAATCCGTCAGAAACAA..... 3174
|||||
817 ThrLeuAlaLeuAsnAsnSerAlaGluLysArgAlaAlaAlaSerValAs 833
|||||
3175 .....GAGCTTTCCGACAACTCGCAAGCGGAGAGAAACAGAGCGCGCT 3220
|||||
833 nTyrAlaLeuGlyGlyAsnLeuThrAsnAsnGlyThrLeuMetThrGlyM 850
|||||
3221 TCACGGCAAAACAGGCA.....CAACTTCGCCCAACACACAGCGGAA 3264
|||||
850 etSerGlyGlnGlnAlaGlyAsnValLeuValLysGlyAsnTyrHis 866
|||||
3265 AAAGACAACGCGCAAGCGCTT..... 3285
|||||
867 GlyAsnAsnGlyGlnLeuValMetAsnThrValLeuAsnGlyAspAspSe 883
|||||
3286 .....GACGCGGTGATTCGCGCGGCGCAATGCCACCGCAAAAGGCA 3328
|||||
883 rValThrAspLysLeuValValGluGly.....AspThrSerG 896
|||||
3329 AAAGTGTTCGCAACCGCGCGCGCAGCAGCGGGGAAATGCC..... 3372
|||||
896 lyThrThrAlaValThrValAsnAsnAlaGlyThrGlyAlaLysThr 912
|||||
3373 .....GCACTT.....ATGCAAGCGGAGGAAGAAAGAAACG 3404
|||||
913 LeuAsnGlyIleGluLeuIleHisValAspGlyLysSerGluGlyGluPh 929
|||||
3405 GGTGACGGGATTAAGACACCGCTT..... 3431
|||||
929 eValGlnAlaGlyArgIleValAlaGlyAlaTyrAspTyrThrLeuAlaA 946
|||||
3432 ..GCGCAACACAGCGCAAGCGGAAC.....CCGCGCGGC 3464
|||||
946 rgGlyGlnGlyAlaAsnSerGlyAsnTrpTyrLeuThrSerGlySerAsp 962
|||||
3465 TACACCGCTTCCCGCGCGCGC.....CCGCGCGCGC 3499
|||||
963 SerProGluLeuGlnProGluProAspProMetProAsnProGluProA 979
|||||
3500 GGGATTTCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3549
|||||
979 snProAsnProGluProAsnProAsnProThrProThrProGlyProAsp 995
|||||
3550 CTG.....ATCAGCCCTTATGCCAATAGCGGTTTGTAGTGA 3584
|||||
996 LeuAsnValAspAsnAspLeuArgProGluAlaGlySerTyrIleAlaAs 1012
|||||

```







[illegible][illegible]





```

1195 LeuGlyAsnThrGlyGlyPheValAspLeuGlyThrTyrGluTyr...Va 1210
1216 TATCCGCAAAAGACGGC...GAGTTCCTCCGCTGATAATCCGCTCAAG 3169
1210 lLeuLysSerAspGlyAsnSerAsnTrpAsnLeuThrAsnAspValLys 1226
3170 AACAGAGCTTTCGGACAAACTCGCAAGCGGGGAGAAACAGAGCGGCC 3219
1226 ..... 1226
3220 TTGACGCAAAACAGGACAACTTCGCGCAAAACACAGCGGAAAAAGA 3269
1227 .....Pro.AsnProAspProAsn..... 1232
3270 CAAGCGCAAAAGCTTGACCGCTGATTGCGCGCGCGCAATGCCACCG 3319
1232 ..... 1232
3320 AAAAGCAGAAAGTGTTCGGAACCGCGCGCGGAGCGCGGGGAAAAAT 3369
1233 .....ProAsn..... 1234
3370 GCGGCGATTATGAGCGCGGAGGAGAGAAAAACGGTGCAGGCGGATAA 3419
1234 ..... 1234
3420 AGACACCGCTTGGCGAAACAGCGGAGGAAACCGCGCGGTACCA 3469
1235 .....ProAsnProAsnPro 1239
3470 CCGCTTCCCGCGCGCGCG...CCGCGCGCGCGGATTTGGCGCAACGG 3516
1240 LysProAsp.ProLysProAspProLysProAsp.ProLysProAspPro 1255
3517 CAGCCCAACACCGCAACCGCGCGCGGACCTGATCAGCGCTTATGCG 3566
1256 ThrProGluProThrProThrProValProGluLysArgIleThrProSe 1272
3567 CAATAGCGGTTGAGTGAATTTTCGCGACGCTCAACAGCGTTTC... 3612
1272 rThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuValPheAspA 1289
3613 .....GCGTACAGGACGAATTTGCGGCTGTTTGGCGAAGAC 3651
1289 laGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAlaSerPro 1305
3652 CGCGCGACCGCTTGGACAGCGGATCCGGGACACCAACACTACCG 3701
1306 HisAsnAsnValTrp...GlyAlaThrTyrAsnThrArgAsnAsnVa 1321
3702 TTCGCAAGATTTCGCGCTACCGGCAACAAACCGACCTGCGCCCAATCG 3751
1321 lThrThrAspAlaGlyAlaGlyPheGluGluThr.....LeuThrG 1335
3752 GTATCGCAAAACACTC..... 3777
1335 lMetThrValGlyIleAspSerProAsnAspIleProGluGlyIleAla 1351
3778 CGCGTCGCATCTTG.....TTTCGCAACCGGACCGGAAACACTT 3821
1352 ThrLeuGlyAlaPheMetGlyTyrSerHisSerHisIleGly.....Ph 1366
3822 CGACGAGCGCATCGGCAACTCGCGACGCTTGCGCACGCTGCGCTTTTCG 3871
1366 eAspArgGly.....GlyHisGlySerVal...G 1375
3872 GCAATACGCGCATCGCG..... 3888
1375 lYSerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisGluSerGlyPhe 1391
3889 .....AGTTCGACATCGGATCAGC... 3909

```

```

1392 TyrLeuAspGlyValValValLysLeuAsnArgPheGluSerAsnValAlaGI 1408
3910 .....GCGGGCGGGGTTTGTAGTAGCGGACGCTT...TCAGACGGCA 3949
1408 yLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrHisSerAsnGlyL 1425
3950 TCAGAGGCAAAATCCGCCCGCTGCTGCATTACGGCATTACGCAAGA 3999
1425 euGlyGlyHis.....IleGluThrGly 1432
4000 TACCOCGAGCTTTCGCGGATTCGGCATCGAAGCGCACATCGCGCAAC 4049
1433 MetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerLeuTh 1449
4050 GCGCTATTTCGTCCAAAAGCGGATTACCGA.....TACGAAA 4087
1449 rGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGlyMetGluS 1466
4088 ACGTCAATATCGCACCGCGGCTTGCATTCAACCGCTACCGC...GCG 4134
1466 erLysSerValAspThrArgSerIle.....TyrArgGluLeu 1478
4135 GGCATTAAAGCAGATTATTCATCAACCGCGCGCAACACATTTCCATCAC 4184
1479 GlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleGI 1495
4185 GCCTATTTCGAGCTGCTTATACCGATCGCGCTTCGCGCAAGTCCGAA 4234
1495 uProTrpLeuLysAlaAlaValArgLysGluPheValAspAsnArgV 1512
4235 CGCGCGCTCAATACG..... 4248
1512 alLysValAsnAsnAspGlyAsnPheValAsnAspLeuSerGlyArgArg 1528
4249 .....GCCGATTTCGCGGAGGATTCGGCAAAACCGCGAGTGC 4286
1529 GlyIleTyrGluAlaGlyIleLysAlaSerPheSerSerThrLeuSerGI 1545
4287 GGAATGGGGGTAAACCGCGCAATCAAAAGGTTTCACGCTGCTCCCTCCACG 4336
1545 yHisLeuGlyVal.....GlyTyrSer.....HisG 1554
4337 CTGCGCGCGCAAGGGCGCG 4356
1554 lYAlaGlyValGluSerPro 1560

```

seq\_name: pir2:D85912

seq\_documentation\_block:

hypoetical protein ypfA [imported] - Escherichia coli (strain O157:H7, substrain ED  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: D85912  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Status: preliminary  
A:Residues: 1-1528 <STO>  
A:Molecule type: DNA  
A:Cross-references: GB:AF005174; NID:g12517083; PIDN:AAG57760.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ypfA

alignment\_scores:

Quality:	344.50	Length:	1543
Ratio:	0.480	Gaps:	84
Percent Similarity:	46.533	Percent Identity:	20.609

alignment\_block:

US-09-303-518d-653 x D85912 ..

Align seg 1/1 to: D85912 from: 1 to: 1528

```

343 CATACGGGGGTATTAACAATGTTGATTTTGGTGGGAGGGAACAATCC 392
: : : : : : : : : : : : : : : : : : : : : : : : : :
265 TyrSerGlyGlyValLeuAspValSer...GlyGlyThrAlaThrAsnVa 280
: : : : : : : : : : : : : : : : : : : : : : : : : :
393 CGATCAGCAGCGCTTTCTTACCAATGTTGAAAAGAAATTAATATAAG 442
||||| : : : : : : : : : : : : : : : : : : : :
280 lThrGlnHis.....AspGlyAlaIleuIysThrAsnGlyTf 295
: : : : : : : : : : : : : : : : : : : : : : : : : :
443 CAGGAGTAACGGCCTCTTATGGCGGATTAATCATATGCCGCTTGG 492
||||| : : : : : : : : : : : : : : : : : : : :
295 hrThrValSerGlyThrAsnSerGluGlyAlaPhe.....SerIle 308
: : : : : : : : : : : : : : : : : : : : : : : : : :
493 CACAATTTGTTCACAGATGCAGAACCTGTTGAGATGACCATTTATGGA 542
||||| : : : : : : : : : : : : : : : : : : : :
309 HisAsnHisValAlaAspAsnValLeuLeuGluAsnGlyGlyHisLeuAs 325
: : : : : : : : : : : : : : : : : : : : : : : : : :
543 TGGGTGGAATACGCTGATTTAAATAAATACCTGATC..... 580
||||| : : : : : : : : : : : : : : : : : : : :
325 pileAsnAlaTyrGlySerAlaAsnIys.ThrIleIleIysAspIysGly 341
: : : : : : : : : : : : : : : : : : : : : : : : : :
581 .....GTGTCGATCGGAGCAGGAGCAGAC.....AATAT 609
||||| : : : : : : : : : : : : : : : : : : : :
342 ThrMetSerValIleThrAsnAlaIysAlaAspAlaThrArgIleAspAs 358
: : : : : : : : : : : : : : : : : : : : : : : : : :
610 TGGCGGTCTGATCAGACAGCAACCCATAACCGGAAAGTTTCATATCATAT 659
||||| : : : : : : : : : : : : : : : : : : : :
358 nGlyGlyValMet...AspValAlaGlyAsn..... 367
: : : : : : : : : : : : : : : : : : : : : : : : : :
660 TGCAACGGCATATCTTGCTGCTGGTGGCAATACCTTTGCACAAAATG 709
||||| : : : : : : : : : : : : : : : : : : : :
368 .....AlaThrAsnThrIleIleAsnGlyGlyThrGlnAsnIleAsnA 382
: : : : : : : : : : : : : : : : : : : : : : : : : :
710 GATCAGGT.....GGTGGCAGCTCACTTAGGTAGCCGCAAAAATTT 750
: : : : : : : : : : : : : : : : : : : : : : : : : :
382 snTyrGlyIleAlaThrGlyThrAsnIleAsnSerGlyThrGlnAsnIle 398
: : : : : : : : : : : : : : : : : : : : : : : : : :
751 AAACATAGCCATATGTTGTTTTTACCACAGGAGGC.....TC 788
||||| : : : : : : : : : : : : : : : : : : : :
399 Lys.....SerGlyGlyLysAlaAspThrTh 407
: : : : : : : : : : : : : : : : : : : : : : : : : :
789 ATTGGCGCAGTGGTCCACCAATGTTTATCTATGATGCCCAAAACCAA 838
: : : : : : : : : : : : : : : : : : : : : : : : : :
407 rIleIleSerSerGlySerArgGlnValVal.....GluLysAspG 421
: : : : : : : : : : : : : : : : : : : : : : : : : :
839 AGTGGTTAATTAATGGGTATTCGAACAGGCAACCCCTATATAGGAAA 888
||||| : : : : : : : : : : : : : : : : : : : :
421 lyThrAlaIleGlySerAsnIleSerAlaGlySerIleuIleValTyr 437
: : : : : : : : : : : : : : : : : : : : : : : : : :
889 AGCAATGGCTCCAGCTAGTTCTGAAGATTTGTTCTATGATGAAATCTT 938
: : : : : : : : : : : : : : : : : : : : : : : : : :
438 ThrGlyGlyIle..... 441
: : : : : : : : : : : : : : : : : : : : : : : : : :
939 TGCTGGAGATACCATTCAGTATCTACGAACACATCAAAATGGGAAT 988
: : : : : : : : : : : : : : : : : : : : : : : : : :
442 .....AlaHisGlyValAsnGlnGlu.....ThrGlySerA 452
: : : : : : : : : : : : : : : : : : : : : : : : : :
989 ACTTTTTTAACGCAATAAATAGGCGCAGGA...AAATCGATGCCAAA 1035
: : : : : : : : : : : : : : : : : : : : : : : : : :
452 laLeuVal.....AlaAsnThrGlyAlaGlyThrAspIleGluGlyTyr 466
: : : : : : : : : : : : : : : : : : : : : : : : : :
1036 CATAAA.....CACTATTTCTCTACCT.....TATAGATT 1064
: : : : : : : : : : : : : : : : : : : : : : : : : :
467 AsnLysLeuSerHisPheThrIleThrGlyGlyGluAlaAsnTyrValVa 483
: : : : : : : : : : : : : : : : : : : : : : : : : :
1065 AAAACACGACCGCTTCAATGTTTATGTTCTTTTATCCGACACAGCAA 1114
: : : : : : : : : : : : : : : : : : : : : : : : : :
483 lIleGluAsnThrGlyGluLeu.....ThrValValAlaLysThrSera 498
: : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

1115 GAGAACCTGTTTATCATGCTGCAGGTGG..... 1143
: : : : : : : : : : : : : : : : : : : :
498 laLysAsnThrThrIleAspAlaGlyGlyLysLeuIleValGlnLysGlu 514
: : : : : : : : : : : : : : : : : : : :
1144 GTCACACAGTTATCGACCCAGACTGAATAATGGAGAAAAATATTCCTTTAT 1193
: : : : : : : : : : : : : : : : : : : :
515 AlalysThrAspSerThrArgLeuAsnAsnGlyValLeuGluValG 531
: : : : : : : : : : : : : : : : : : : :
1194 TGCAAAAGAAAGAGTGAATGTACTACCAACACATCAACCAAGGG 1243
||| : : : : : : : : : : : : : : : : : : : :
531 nAspGlyGlyGluAlalys.....HisValGluGlnGlnS 543
: : : : : : : : : : : : : : : : : : : :
1244 CGGCGGTTGTTATTTTACGGGTAAATTTACGGTCTCGCTCAAAACAAAC 1293
: : : : : : : : : : : : : : : : : : : :
543 erGlyGlyAlaLeuIleAlaSerThrThrSerGlyThrLeuIleGluGly 559
: : : : : : : : : : : : : : : : : : : :
1294 GAACAGTGGCAAGCGCGGCTTCATATCAGTATGATGGCAGTACCGTTAC 1343
: : : : : : : : : : : : : : : : : : : :
560 ThrAsnSerTyrGlyAspAlaPheTyrIleArgAsnSerGlu..... 573
: : : : : : : : : : : : : : : : : : : :
1344 TTGCAAGTAAACGGCGTGGCAACGACCGCTCTCCAAAATCGGCAAG 1393
: : : : : : : : : : : : : : : : : : : :
574 .....AlalysA 576
: : : : : : : : : : : : : : : : : : : :
1394 GCACGCTGCTGTTTCAAGCCAAAGGGGAAACCAAGCTCGCTCAGCGTG 1443
: : : : : : : : : : : : : : : : : : : :
576 snValValLeu.....GluAsnAlaGlySerLeuThrVal 587
: : : : : : : : : : : : : : : : : : : :
1444 GCGACGGTAAAGTCATCTTAGATCAGCAGCGGCGGACGATCAAGGCAAAA 1493
||||| : : : : : : : : : : : : : : : : : : : :
588 ValThrGlySerArgAlaValAspThrIleIleAsnAlaAsnGlyLysMe 604
: : : : : : : : : : : : : : : : : : : :
1494 ACAAGCCTTT...ACTGAAATCGGCTTGGTC.....AGCGCAGGG 1531
: : : : : : : : : : : : : : : : : : : :
604 tAspValTyrGlyLysAspValGlyThrValLeuAsnSerAlaGlyThrG 621
: : : : : : : : : : : : : : : : : : : :
1532 GAGCGGTCAACTGAATGCCGATATCAATCAGTTCAACCCCGCAAACTCTAT 1581
||||| : : : : : : : : : : : : : : : : : : : :
621 lInThrIleTyrAlaSerAlaThrSer.....AspLysAla... 632
: : : : : : : : : : : : : : : : : : : :
1582 TTCGGCTTCCGCGGAGCTTTGGATTTGAACGGGATTCGCTTCGTT 1631
: : : : : : : : : : : : : : : : : : : :
633 ...AsnIleLysGlyGlyLysGlnThrValTyrGlyLeuAlaThrGlu... 647
: : : : : : : : : : : : : : : : : : : :
1632 CCACCGCATTCAAAATACCGATGAAGGGCGGATGTTGTCAACCAACATC 1681
||| : : : : : : : : : : : : : : : : : : : :
648 .....AlaAsnIleGluSerGlyGlnIleValAspGlyGlys 661
: : : : : : : : : : : : : : : : : : : :
1682 AAGACAAAGATCCACCGTTACCATTCAGGCAATAAGATATTACTACA 1731
: : : : : : : : : : : : : : : : : : : :
661 erThrGluLysThrHisIleAsn...GlyGlyThrGlnThrValGlnAsn 676
: : : : : : : : : : : : : : : : : : : :
1732 ACCGCAATACCAACACTTGGAT.....AGCAAAAAGAAATTCG 1772
||||| : : : : : : : : : : : : : : : : : : : :
677 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnIleMe 693
: : : : : : : : : : : : : : : : : : : :
1773 CTACACGCTTGGTTTGGGAGAAAGATCCAAACAAACGACGCGCGCG 1822
||| : : : : : : : : : : : : : : : : : : : :
693 tAlaAsn.....GlyThrAlaGluGlySerIleIleAsnGlyGlys 707
: : : : : : : : : : : : : : : : : : : :
1823 TCATCTGAATTAACCAACCGGAGAGCGGATCGCATTTACTGCTTCC 1872
: : : : : : : : : : : : : : : : : : : :
707 erGlnIleValAsnGluGlyGlyLeuAlaGluAsnSerValLeuAsnAsp 723
: : : : : : : : : : : : : : : : : : : :
1873 GCGGAACA.....AATTTAAACGCAATATCAC 1901
||||| : : : : : : : : : : : : : : : : : : : :
724 GlyGlyThrLeuAspValArgGluLysGlySerAlaThrGlyIleGlnGl 740
: : : : : : : : : : : : : : : : : : : :
1902 GCAACAAACGCAACACTGTTTTTTCAGCGGCAACGACGACCGCGCCT 1951
: : : : : : : : : : : : : : : : : : : :
740 nSerSerGlnGlyAlaLeuValAlaThrThrArgAlaThrArgValThr 756
: : : : : : : : : : : : : : : : : : : :
1952 ACATCATTTAGCAACGCGGTGTCAAAATGCAAGGTATCCCAAGGA 2001
: : : : : : : : : : : : : : : : : : : :

```





4249 .....GCGGTATTGGCGAGGATTTCGCAAAACCCGAGTGC 4286  
 1486 GlyIleTyrGlnAlaGlyIleLysAlaSerPheSerThrLeuSerCl 1502  
 4287 GGAATGGGCGTAACGCCGAAATCAAGGTTTCACGGTGTCCCTCCAGC 4336  
 1502 yHisLeuGlyVal.....GlyTyrSer.....HisG 1511  
 4337 CTGCCCGCCGCAAGGGCGG 4356  
 1511 lyAlaGlyValGluSerPro 1517

seq\_name: pir2:E95206

seq\_documentation\_block:  
 cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain  
 C:Species: Streptococcus pneumoniae  
 C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
 C:Accession: E95206  
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Morris-  
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,  
 nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris-  
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A:Reference number: A95000; MUID:21357209; PMID:11463916  
 A:Accession: E95206  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-4776 <KUR>  
 A:Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:gl4973269; GSPDB:GN00164; TIGR:  
 A:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SP1772

alignment\_scores:  
 Quality: 329.50 Length: 994  
 Ratio: 0.586 Gaps: 27  
 Percent Similarity: 56.539 Percent Identity: 20.724

alignment\_block:  
 US-09-303-518D-653 x E95206 ..  
 Align seg 1/1 to: E95206 from: 1 to: 4776

1501 TTTAGTGAATCGGCTTGTTCAGCGGACGGGACGGTGCACACTGAATGC 1550  
 186 PheSerGlyAsnThrIleValAsnGlyAlaProAlaIleAsnAlaSerLe 202  
 1551 CGATAATCAGTTCAACCCCGACAACTCTATTTCGGCTTCGCGGGGAC 1600  
 202 uAsnIleAlaLysSerGluThrLysValTyr.....ThrGlyGluG 216  
 1601 GTTTCGATTTCACGGGCGATTTCGCTTTTCACCGCATTCACAAATACC 1650  
 216 lyValAspSerValTyrArgValProIleTyrLysLeuLysValThr 232  
 1651 GATGAAGGGGCGGATGATTGTCAACCAATCAAGACAAAGAAATCCACGT 1700  
 233 AsnAspGlySerLysLeuThr.....PheThrTy 242  
 1701 TACCAATTACA...GCCAATAAGATATTACTACACCGGCAAT..... 1740  
 242 rThrValThrTyrValAsnProLysThrAsnAspLeuGlyAsnIleSerS 259  
 1741 .....AACACAACTTGGATAGCAAAAAA 1764  
 259 erMetArgProGlyTyrSerIleTyrAsnSerGlyThrSerThrGlnThr 275  
 1765 GAAATTCCTACACGGTGTGTTGGCGAAGATGCAACCAACG... 1812  
 276 MetLeuThrLeuGlySerAspLeuGlyLysProSerGlyValLysAsnTy 292

3517 CAGCCCAACCGCAACCCCAACCGCAGCGACCTGTATCATCCGCTTATGC 3566  
 1213 ThrProGluProThrProValProGluLysArgIleThrProSe 1229  
 3567 CAATAGCGGTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTC.... 3612  
 1229 rThrAlaAlaValLeuAsnMetAlaAlaThrLeuProLeuValPheAspA 1246  
 3613 .....GCGGTACAGCACGAATTTGGACCGGCTTTCGCCGAAGAC 3651  
 1246 laGluLeuAsnSerIleArgGluArgLeuAsnIleMetLysAlaSerPro 1262  
 3652 CGCCCAACCGCGTTTGGCAAGCGGATCCGGACACCAACACTACCG 3701  
 1263 HisAsnAsnValTrp...GlyAlaThrTyrAsnThrArgAsnAsnVa 1278  
 3702 TTCGCAAGATTTCGCGCTACCGCCCAACACCGACCTGGCCAAATCG 3751  
 1278 lThrThrAspAlaGlyAlaGlyPheGluGlnThr.....LeuThrG 1292  
 3752 GTATCGAGAAAACCTC.....GGCAGCGGG 3777  
 1292 lyMetThrValGlyIleAspSerProAsnAspIleProGluGlyIleAla 1308  
 3778 CGCGTCGCATCTG.....TTTTCGCACACCGCGCCGGAACACCTT 3821  
 1309 ThrLeuGlyAlaPheMetGlyTyrSerHisHisIleGly.....Ph 1323  
 3822 CGACGACGCGATCGCAACTCGCGCGCTTCCCGACGCTGCGGTTTCG 3871  
 1323 eAspArgGly.....GlyHisGlySerVal...G 1332  
 3872 GGCAATACGGCATCGCGC..... 3888  
 1332 lySerTyrSerLeuGlyTyrAlaSerTrpGluHisGluSerGlyPhe 1348  
 3889 .....AGGTTTCGACATCGGCATCAGC... 3909  
 1349 TyrLeuAspGlyValValLysLeuAsnArgPheGluSerValAlaGl 1365  
 3910 .....GCGGCGCGGGTTTGTAGTAGCGGACGCTT...TCAGACGGCA 3949  
 1365 yLysMetSerSerGlyAlaAlaAsnGlySerTyrHisSerAsnGlyL 1382  
 3950 TCAGAGGCAAAATCCCGCGCGTCTCATTCACGGCATTCAGGCAAGA 3999  
 1382 euGlyGlyHis.....IleGluThrGly 1389  
 4000 TACCGCGCAGGTTTCGCGGATTCGGCATCGAACCGCACATCGGCGCAAC 4049  
 1390 MetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSerLeuTh 1406  
 4050 GCGCTATTTCGTCACAAACCGGATTACCGA.....TACGAAA 4087  
 1406 rGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGlyMetGluS 1423  
 4088 ACCTCAATATCGCACCGCGGCTTGCATTCAACCGCTACCGC...GCG 4134  
 1423 erLysSerValAspThrArgSerIle.....TyrArgGluLeu 1435  
 4135 GGCATTATGAGCAGATTATTCATTCAACCGCGCGCACATATTCATCAC 4184  
 1436 GlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGluIleGl 1452  
 4185 GCCTTATTGAGCGTGTTCATACCGATCGCGTTCGCGCAAGTCCGAA 4234  
 1452 uProTrpLeuLysAlaAlaValArgLysGluPheValAspAsnArgv 1469  
 4235 CGCGCGTCAATACC..... 4248  
 1469 aiLysValAsnAsnAspGlyAsnPheValAsnAspLeuSerGlyArgArg 1485

1813 .....AACGGCGGCTCAATCTGAATTACCAACCGGAAG 1849  
292 rIleThrAspLysAsnGlyArgGlnValLeuSerTyrAsnThrSer... 307  
1850 CGGATCGCAGCTTTACTGCTTCCGGCGGAACAAATTTAAACGGCAATATC 1899  
308 .....ThrMetThrThrGlnGlySerGlyTyrThrTrpGlyAsnGly 321  
1900 ACGCAACAACAGCGCAACTGTTTTTCAGGGCGAGACCGACCGCACGC 1949  
322 AlaGlnMetAsnGly.....PhePheAlaLysLys.....GI 332  
1950 CTACAATCATTTAGAACGGCGGTGCTCAAAATGGAAGGTATCCCA...C 1996  
332 yTyrgly...LeuThrSerSerTrpThr.....ValProIle 344  
1997 AAGGAGAAATCGTGTGGGCAACAGATTTGGATCGACCGCACATTTAAAGC 2046  
344 hrGlyThr.....AspThrSerPheThrPheThrProTyrAlaAla 357  
2047 GAA.....AACTTCCATPATTCAGGGCGGACAAACGGT 2078  
358 ArgThrAspArgIleGlyIleAsnTyrPheAsnGlyGlyGlyValVa 374  
2079 GGTTCCTCCGCAATGTTCCCAAGATGGAACGGATTTGGCATTTAAGCAATC 2128  
374 lGluSerThrThrSerGln.....SerLeuSerGlnS 386  
2129 ACGCCCAAGCAGTTTTCGTGTGTCGCCACCGCATCAAAAGCCACACATCTGT 2178  
386 erLysSerLeuSerValSerAlaSerGlnSerAlaSerAlaSerAlaSer 402  
2179 ACAGGTCGAGCTGGACGGTCTGACAACTTGTACCGAAAAACCATATC 2228  
403 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerTh 419  
2229 CGACGATAAAGTGATTCTTCATTAGCAAGACCGACATCAGAGCAATG 2278  
419 rSerAlaSerAlaSerAlaSerThrSer.....AlaServ 431  
2279 TCAGCCTTCGGATCAGCTCATTTAAATCTCAGAGGACTTGCCACATC 2328  
431 aSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAla 447  
2329 AACGGCAATCTTAGTCAGCGGAGACACGACCTATAGCTTACCGCAA 2378  
448 SerAlaSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlase 464  
2379 CGCCACCCCAACCGCAACCTCAGCTCGTGGCAATGCCAAGCAACAT 2428  
464 rAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerA 481  
2429 TTAAT...CAAGCCACATTAACGGCAACATATCGGCTTCGGACAATGCT 2475  
481 laSerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAla 497  
2476 TCATTTAATCTAAGCAACACCGCGCTACAAAACGCGACTGTGAGCTTTC 2525  
498 SerThrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAlase 514  
2526 CGAACACGCTAAGCAACAGTAAGCAATTCGCGACTCAACGGAATGCT 2575  
514 rThrSerAlaSerAlaSerAlaSerThrSerAla...SerGlySerAlas 530  
2576 CCTAGCGGTAAGGACGATTTCCATTTTGAACACAGCCGCTTTACCGGA 2625  
530 erThrSerThrSerAlaSerAlaSerThrSerAlaSerAla...SerAla 545  
2626 AAATCATCGCGCGCAAGGATACGGCATTTACCTTAAAGACACGCGAATG 2675  
546 SerThrSerAlaSerAlaSerAlaSerIleSerAlaSerGluSerAlase 562

2676 GACGCTCGCTGGCGCAGCGAATTTAGGCAATTTAAACCTTTGACAACGCCA 2725  
562 rThrSerAlaSerGluSerAlaSerThrSerThrSerAlaSerAlaSerT 579  
2726 CCATTACACTCAATTCGCTATCGACACGATGCGCGCGCGCAACACC 2775  
579 hrSerAlaSerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThr 595  
2776 GGCAGTGGCGGAGATGCGCGCGCGCGCTTCCGCGCTTCCCTATTATC 2825  
596 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSe 612  
2826 COTTTACGCGCGCAACTTCGGCA...GAATCCGTTTCAACACGCTGACGG 2872  
612 rAlaSerAlaSerThrSerAlaSerGluSer...AlaSerThr..... 625  
2873 TAAACGGCAATTTGAACGCTCAGGGAACATTCGCTTTATGTCGGAACATC 2922  
626 ...SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerTh 641  
2923 TTCGGCTACCGCAGCGGCAAAATTTGAAGCTGGGGAAGTTCCGAAGGCAC 2972  
641 rSerAlaSerAlaSerAlaSerThrSer..... 650  
2973 TTACACCTTGCTGTCAACAATACCGCAACGACCGCTAAGTCTCGAGC 3022  
650 ..... 650  
3023 AATTGACGGTAGTGAAGGAAAGACACACACCGCTGTCCGAAAAATCTT 3072  
651 .....Al 651  
3073 AATTTCACCTTCGCAAAACGAAACAGCTCGATCCGCGCATGGCGTTATCA 3122  
651 aSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlas 668  
3123 GCTTATCCGAAAGACGCGGAGTTCCGCTGCATTAATCCGCTCAAGAAGC 3172  
668 erAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSer 684  
3173 AAGAGCTTTCGCAAACTCGGCAAG.....CGGGGAAACACAGA 3212  
685 GluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAl 701  
3213 GGCGGCTTGACGCAAAACAGCGCAACTTGCAGCAACACACACAGCGCG 3262  
701 aSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlas 718  
3263 AAAAGACAAACGCGCAAGCCTTTCAGCGCTGATTCGCGCGCGCGCAAT 3312  
718 erAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSer 734  
3313 GCCACGAAAGCAGAAAGTGTTCGCAACCGCGCGCGCGCGCGCGCGCG 3362  
735 AlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAl 751  
3363 GGAA.....AATCGCGCTATTATGACGCGGAGAGAGAGAAAAAAC 3403  
751 aSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlas 768  
3404 GGGTGCGAGCGGATAAGACACCGCTTTCGCAACACAGCGCGCGAGGAA 3453  
768 erThrSerAlaSerAlaSerAlaSerIleSerAlaSerGluSerAlaSer 784  
3454 ACCGCGCGGTACACCGCTTTCGCGCGCGCGCGCGCGCGCGCGGGA 3503  
785 ThrSerAlaSerAlaSerAla...SerThrSerAlaSerAlaSerAlaSerT 801  
3504 TTTGCGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGA 3553  
801 hrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThr 817  
3554 TCAGCGCTTATGCAATAGCGGTTTGTAGTGATTTTCCGCCACGCTCAAC 3603

```
C;Accession: T08658
R:Plunkett, G.
submitted to the EMBL Data Library, September 1995
A:Reference number: Z16465
A:Accession: T08658
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-949 <PLU>
A:Cross-references: EMBL:U36840; NID:g10331110; PID:g1033141
A:Experimental source: strain K12, substrain MG1655
C:Genetics:
A:Map position: 57 min

alignment_scores:
  Quality: 295.50      Length: 1157
  Ratio: 0.554        Gaps: 57
  Percent Similarity: 46.067  Percent Identity: 20.138

alignment_block:
US-09-303-518D-653 x T08658 ..
Align seg 1/1 to: T08658 from: 1 to: 949

1420 GAAACCAACAGGCTCGTCAAGGCGGCGGTAAGTCATCTTAGATCA 1469
||||| ||||||| ||||||| ||||||| |||||||
3 GluAsnAlaGlySerLeuThrValThrGlySerArgAlaValAspTh 19

1470 GCAGGCGGACGATCAAGCAAAACAAAGCCTTT...AGTGAATCGGCT 1516
||||| ||||||| ||||||| ||||||| |||||||
19 rIleIleAsnAlaAsnGlyLysMetAspValTyrGlyLysAspValGly 36

1517 TGCTC.....AGCGGCGGCGGACGTGCAACTGAATGCCGATAAT 1557
||||| ||||||| ||||||| ||||||| |||||||
36 hrValLeuAsnSerAlaGlyThrGlnThrIleTyrAlaSerAlaThrSer 52

1558 CAGTTCAACCCCGACAAACTCTATTTTCGGCTTCGCGGCGGAGCTTGGGA 1607
||||| ||||||| ||||||| ||||||| |||||||
53 .....AspLysAla.....AsnIleLysGlyGlyLysGlnTh 63

1608 TTTGAACGGGCATTCGCTTTTCGTTCCACCGCATTCACAAATACCGATGAAG 1657
||||| ||||||| ||||||| ||||||| |||||||
63 rValTyrGlyLeuAlaThrGlu.....AlaAsnIleGluSerG 76

1658 GGGCGCATGATTGTCAACACACATCAAGACAAAGAAATCCACCGTTACCAT 1707
||||| ||||||| ||||||| ||||||| |||||||
76 lyGluGlnIleValAspGlyGlySerThrGluLysThrHisIleAsn... 91

1708 ACAGGCATAAAGATATTACTACACCGGCAATAACAACAACTTGGAT.. 1755
||||| ||||||| ||||||| ||||||| |||||||
92 GlyGlyThrGlnThrValGlnAsnTyrGlyLysAlaIleAsnThrAspIl 108

1756 .....AGCAAAAAGAAATTCGCTACACGCTTGGTTGGCGAGAAAG 1798
||||| ||||||| ||||||| ||||||| |||||||
108 eValSerGlyLeuGlnIleMetAlaAsn.....GlyThrAlaG 122

1799 ATGCAACAAACAAACGACGCGGCTCAATCTGAATTCACCAACCGAAGAA 1848
||||| ||||||| ||||||| ||||||| |||||||
122 luGlySerIleIleAsnGlyGlySerGlnValValAsnGluGlyLeu 138

1849 GCGGATCCCACTTTACTCTTTCGCGGCGGAACA..... 1881
||||| ||||||| ||||||| ||||||| |||||||
139 AlaGluAsnSerValLeuAsnAspGlyGlyThrLeuAspValArgGluLy 155

1882 ....AATTTAAACGGCAATATCACGCAAAACAAACGCAAACTGTTTTC 1927
||||| ||||||| ||||||| ||||||| |||||||
155 sclySerAlaThrGlyIleGlnGlnSerSerGlnGlyAlaLeuValAla 172

1928 GCGGCAGACCGACACCGCCTTACATCATTTAGGAAGCGGGTGGTCA 1977
||||| ||||||| ||||||| ||||||| |||||||
172 hrThrArgAlaThrArgValThr.....GlyThr 181

1978 AAAATGGAGGTATCCCAACAGGAGAAATCGTGTGGGACACGATGGAT 2027
||||| ||||||| ||||||| ||||||| |||||||
```

```
818 SerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSe 834
||||| ||||||| ||||||| ||||||| |||||||
3604 AGCGTTTTCGCGTACAGGACGAATTCGACCGCGTGTTCGCCGAAGACCG 3653
||||| ||||||| ||||||| ||||||| |||||||
834 rAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerA 851
||||| ||||||| ||||||| ||||||| |||||||
3654 CCGCAACCGCGTTGGACACGCGCATCCGGGACACCAACACACTACCGTT 3703
||||| ||||||| ||||||| ||||||| |||||||
851 laSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAla 867
||||| ||||||| ||||||| ||||||| |||||||
3704 CGCAAGATTTCGCGCGCTACCGCAACAAACCGCACCTCGCCCAATCGGT 3753
||||| ||||||| ||||||| ||||||| |||||||
868 SerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSe 884
||||| ||||||| ||||||| ||||||| |||||||
3754 ATGCAGAAAAACCTCGGACGCGGCGGTGCGCATCTCTGTTTCGCACAA 3803
||||| ||||||| ||||||| ||||||| |||||||
884 rGluSerAlaSerThrSerAlaSerAlaSerAlaSer.....ThrS 898
||||| ||||||| ||||||| ||||||| |||||||
3804 CCGGACCGGAACACCTTCGACGACGCGCATCGCAACTCGGCACGCGTTG 3853
||||| ||||||| ||||||| ||||||| |||||||
898 erAlaSerAlaSerThrSerAlaSerAlaSerAlaSerAlaSerThrSer 914
||||| ||||||| ||||||| ||||||| |||||||
3854 CCCACGGTCCGCTTTCGCGCAATACGCGCATCGGCGGTTCGACATCGGC 3903
||||| ||||||| ||||||| ||||||| |||||||
915 AlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAl 931
||||| ||||||| ||||||| ||||||| |||||||
3904 ATCAGCGGCGGCGGCTTTAGTAGCGGACGCTTTCAGACGCGCATCAG 3953
||||| ||||||| ||||||| ||||||| |||||||
931 aserAlaSerAlaSerThrSerAlaSerAlaSerThrSerAlaSerGluS 948
||||| ||||||| ||||||| ||||||| |||||||
3954 AGGCAAAATCCCGCCGCGCTGCTCAATTCAGGCAATTCAGGCAAGATACC 4003
||||| ||||||| ||||||| ||||||| |||||||
948 erAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSer 964
||||| ||||||| ||||||| ||||||| |||||||
4004 GCGCAGGTTTCGCGGATTCGCGCATCGAACCGCATCGGCGCAACGCGC 4053
||||| ||||||| ||||||| ||||||| |||||||
965 AlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAl 981
||||| ||||||| ||||||| ||||||| |||||||
4054 TATTTCTGTCAAAACCGGATTCAGGATACGAAACGTCATATCGCCAC 4103
||||| ||||||| ||||||| ||||||| |||||||
981 aserThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAla 998
||||| ||||||| ||||||| ||||||| |||||||
4104 CCGCGGCTTCATTCACCGCTACCGCGCGGCTTAAGCGAGATATT 4153
||||| ||||||| ||||||| ||||||| |||||||
998 erThrSerAlaSerAlaSerAlaSerThrSerAla..SerAlaSerAla 1014
||||| ||||||| ||||||| ||||||| |||||||
4154 CATTCAACCGGCGCAACATTCATCAGCGCTTATTTAGCGCTGTCC 4203
||||| ||||||| ||||||| ||||||| |||||||
1014 erThrSerAlaSerAlaSerAlaSerIleSerAlaSerGluSerAlaSer 1030
||||| ||||||| ||||||| ||||||| |||||||
4204 TATACGATCCGCTTCGCGCAAGTCCGACCGCGTCAATACCGCGCT 4253
||||| ||||||| ||||||| ||||||| |||||||
1031 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerValSerAlaSerTh 1047
||||| ||||||| ||||||| ||||||| |||||||
4254 ATTGGCGCAGGATTTCGCAAAACCGCGAGTGGGAATGGCGGTAACG 4303
||||| ||||||| ||||||| ||||||| |||||||
1047 rSerAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThS 1064
||||| ||||||| ||||||| ||||||| |||||||
4304 CCGAAATCAAAGTTTCAGCGCTGCTCCACGCTCCCGCGCCGCAAGGG 4353
||||| ||||||| ||||||| ||||||| |||||||
1064 erAlaSerAlaSerAlaSerThrSerAlaSerGluSerAlaSerThrSer 1080
||||| ||||||| ||||||| ||||||| |||||||
4354 CCGCAATGGAACGCGACGACGCGG 4380
||||| ||||||| ||||||| ||||||| |||||||
1081 AlaSerAlaSerAlaSerThrSerAla 1089
||||| ||||||| ||||||| ||||||| |||||||
```

seq\_name: pir2:T08658  
seq\_documentation\_block:  
hypothetical protein f949 - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999

182 ArgAlaAspGlyValAla..... 187  
 2028 CGACCGCACATTTAAAGCGGAAACCTTCCATATTTTCAGGGCGGACAACGGC. 2076  
 188 .....PheSerIleGluGlnGlyAlaAla 196  
 2077 .....GTGGTTTCCCGCAATGTTGCCAAAGTGGGAAGCGGATTTGG 2115  
 196 snAsnIleLeuLeuAlaAsnGlyGlyValLeuThrValGluSerAspThr 212  
 2116 CATTTTAAGCAATCACGCCCAAGCAGTTTCGGTGTCCGCACCGCATCAAG 2165  
 213 Ser.....SerAspLysThrGlnValAsnMetGly.....GlyAr 224  
 2166 CCACACATCTGTACAGGTTCCGAGTGGAGCGGTCTGACAAAGTTGTACCG 2215  
 224 gGluIleValLysThrLysAlaThrAlaThrGlyThrLeuThrLeuGlyG 241  
 2216 AAAAACAACCATCCGACGATAAAGTGATTCCTTCATTGAGCAAGACCCGAC 2265  
 241 LyGluGlnIleValGlu..... 246  
 2266 ATCAGAGGCAATGTCCAGCTTCCGATCACGCTCATTTAAATCTCACAGG 2315  
 247 .....GlyValAlaAsnGluThrThrIleAsnAspGlyG 258  
 2316 ACTTGGCCACATCAACGGCAAT.....CTTAGTG 2344  
 258 yIleGlnThrValSerAlaAsnGlyGluAlaIleLysThrLysIleAsnG 275  
 2345 CAGGGGAGACGCGACTATACGTTAGCGCAACGCCACCCCAAAACGCC 2394  
 275 LuGlyGly.....ThrLeuThrValAsn.....AspAsnGly 285  
 2395 AACCTCAGC...CTCGTGGCAATGCCCAACGACACATTTAATCAAGCCAC 2441  
 286 LysAlaThrAspIleValGlnAsnSerGlyAlaAlaLeuGlnThrSerTh 302  
 2442 ATTTAAACGGC.....AACCATCGGCTTCGACAATGCTTCATTTAATC 2485  
 302 rAlaAsnGlyIleGluIleSerGlyThrHisGlnTyGlyThrPheSerI 319  
 2486 TAAGCAACAACGGGTACAAAC.....GGCAGTCTG 2517  
 319 leSerGlyAsnLeuAlaIaThrAsnMetLeuLeuGluAsnGlyGlyAsnLeu 335  
 2518 ACGGTTTCCGACACGCTAAGGCAAAACGTAAGCATTCCTCCGACCTCAACGG 2567  
 336 LeuValLeuAlaGlyThrGluAlaArgAspSerThrValGlyLysGlyG 352  
 2568 CAATGTCTCCTAGCCGATAGGCGATTTCCATTTTGAACACAGCCGCT 2617  
 352 y.....AlaMetGlnAsnLeuGlyGlnAspSerA 362  
 2618 TTACCGGAAAAATCAGCGCGCGCAAGGATACGCAATTACACTTTAAAGAC 2667  
 362 laThrLysValAsnSerGlyGlyGlnTyThrLeuGlyArgSerLysAsp 378  
 2668 AGCGAATGACGCTCGCGTGGCGACCGGAATTA..... 2700  
 379 GluPheGlnAlaLeuAlaArgAlaGluAspLeuGlnValaIaGlyGlyTh 395  
 2700 ..... 2700  
 395 rAlaIleValTyAlaGlyThrLeuAlaAspAlaSerValSerGlyAlaT 412  
 2701 .....GGCAATTTAAACCTT..... 2715  
 412 hrGlySerLeuSerLeuMetThrProArgAspAsnValThrProValLys 428  
 2716 .....GACACGCCGACCATTTACATTC 2741  
 187 .....

429 LeuGluGlyAlaValArgIleThrAspSerAlaThrLeuThrLeuGlyAs 445  
2742 CGCCTAT .....CGACACATCGCAGCGCGCAACCGGAGT. 2781  
::: ||| :::::::::::::: |  
445 nGlyValAspThrThrLeuAlaAspLeuThrAlaAlaSerArgGlySerV 462  
2782 .....CGGCAGATCGCGCCGCCGGT 2805  
|||::: :  
462 alTrpLeuAsnSerAsnAsnSerCysAlaGlyThrSerAsnCysGluTyrr 478  
2806 TCGCGCGGTCCCTATT .....TC 2825  
:::|||||||  
479 ArgValAsnSerLeuLeuLeuAsnAspGlyAspValrLeuSerAlaGI 495  
2826 COTTACGCCGCCAACTTCGGCAGAATCCCGTTTCAACACGCTCACGATAA 2875  
:::::: |  
495 nThrAlaAlaProAlaThrThrAsnGlyIleItyrAsnThrLeuThrThRA 512  
2876 ACGGCAAATTGAACGTCAGGCAACAATTCGCTTTATGTCGGAACCTCTTC 2925  
|| :::::::::::::::  
512 sn...GluLeuSerGlySerGlyAsnPheTyrlLeuHisThrAsnValAla 527  
2926 GGCTACCAGCGGCAAAATGAAGTCGGCGAAAGTTCGGAAGCACTTA 2975  
||| |||::: |  
528 GlySerArgGlyAspGlnLeuValValAsnAsnAsnAlaThrGlyAsnPh 544  
2976 CACTTGCTCTCAACATACCGGCAACGACCCGTAAGTCTCGAGCAAT 3025  
::: |  
544 eLysIlePheValGlnAspThrGlyValSerProGlnSerAspAspAlaM 561  
3026 TGACGGTAGTGAA.....GGAAAAGACAACACACCGCTGTCCGAAAAAT 3069  
561 etThrLeuValLysThrGlyGlyAsp..... 570  
3070 CTTAATTTCACCTGCAAAACGAA.....CACGTGCGATCGCGGCATG 3113  
:::|||  
571 AlaSerPheThrLeuGlyAsnThrGlyGlyPheValAspLeuGlyThrTy 587  
3114 GCGFTATCAGTTATCGCGAAAGACGGCGAGTTCGCGCTGCATAATCCGG 3163  
587 rGluTyrr...ValLeuLysSerAspGly.AsnSer..... 597  
3164 TCRAAGAACAGAGCTTTCGCACAACTCGCAAGCGGGAGAACACAG 3213  
597 ..... 597  
3214 GCGCGCTTGACGCAAAACAGGCACAACTTGCGCCCAACACACAGCGGA 3263  
597 ..... 597  
3264 AAAGACACGCGCAAGCCCTGACGGCTGATTCGGCGCGCGCGCAATG 3313  
598 .....AsnTrpAsnLeuThrAsnAspValLysProAsnProAsp 610  
3314 CCACCGAAAAGGAGAAAGTTGTCGCGAACCGCCCGCGGACGAGCGCGG 3363  
||| Pro.....IleProAsn..... 614  
3364 GAATATGCGCGCATATGAGCGCGGAGGAAGAGAAAAACGGGTGACGC 3413  
614 ..... 614  
3414 GGATAAAGACACCGCTTGGCGAAACAGCGCAGCGGAAACCCGGCGG 3463  
615 .....ProLysProAspPro 619  
3464 CTACACCGCCTTCCCOCGCGCGCGCGCGCGCGGATTCGCGCAA 3513  
620 LysProAspProLysProAspPro..AsnProLysProAsp...ProThr 634  
3514 CGCGAGCCCCACCGCAACCCCAACCGCAGCGCGACCTGATCAGCGGTTA 3563  
635 ProAspProThrProThrProValProGluLysArgIleThrPro.... 649



[illegible]

405 eValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu.....A 418  
1679 ATCAAGACAAAGAATCCACGGTTACC..... 1704  
418 snIleAspAlaGlyGlyAsnValThrValAsnGlnGlySerPheAlaGly 434  
1705 ...ATTACAGGCAATAAGATATTACTACACCGCCAAATAACAAACAAC.. 1749  
435 IleIleGluGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVa 451  
1750 .TTGATAGCAAAAAGAAATTCCTACACCGTGTGGTGGCGAGAAAG 1798  
451 lleuAlaGlyAlaGlnSerMetAlaLeuThrGly.....A 463  
1799 ATGCACCAAAACGAGCGCGCTCAATCTCAATTACCAACCGGAAGA 1848  
463 spIleValValAspAspGlyAlaVal...LeuSerLeuGluGlyAspAla 478  
1849 GCGGATCGCATCTTACTGCTTTCGCGCGGAACAAAT.....TTAAACGG 1892  
479 AlaAspLeuThrAlaLeuGlnAspAspProGlnSerIleValLeuAsnG 495  
1893 CAATATCATCGCAACAAACGCGCAAACTGTTTTCAGCGCGACACGACAC 1942  
495 yGlyValLeuAspLeuSerSp..... 502  
1943 CGCAGCGCTACAATCATTTAGGAAGCGGTGGTCAAAATGGAAGTATC 1992  
503 .....PheSerThrTrpGlnSerGlyThrSerTyrAsnAspGlyLeu 516  
1993 .....CCACAAGGAGAATCGTGTGGACAACAGATTGGATCGCA 2030  
517 GluValSerGlySerSerGlyThrValIleGlySerGlnAspValValAs 533  
2031 CCGCACATTTAAACCGGAACACTTCCATATTACGGC...GGACAAGCGG 2077  
533 pLeuAla...GlyGlyAspAsnLeuHisIleGlyGlyAspGlyLysAspG 549  
2078 TGGTTTCCCGCAATGTTGCCAAAGTGGAAGCGATTGCGCATTTAAGCAAT 2127  
549 lyValTyrValValValAspAlaSerAspGlyGlnValSerLeuAlaAsn 565  
2128 CAGCCCCAAGCAGTTTTCGTGTGCGACCGCATCAAGGCCACACAATC.. 2175  
566 Asn...AsnSerTyrLeuGlyThrThrGlnIleAlaSerGlyThrLeuMe 581  
2176 .TGTACAGTTCGGACTGGAGCGGTCTGACAACTGTACCAAGTTGTACCCAAAACCA 2224  
581 tValSerAspAsnSerGlnLeuGlyAspThrHisTyrAsnArgGlnValI 598  
2225 TTACCGCAGCATAAA.....GTGATTGCTTCATTGACGACGACC 2262  
598 lePheThrAspLysGlnGlnSerValMetGluIleThrSerAspVal 614  
2263 GACATCAGAGGCAATGTCACG.....CTTGC.....CTTGC 2288  
615 AspThrArgSerAspAlaAlaGlyHisGlyArgAspIleGluMetArgAl 631  
2289 CGATCAGCGTCAATTAATCTCCACAGACTTGCACACTCAACGGCANTC 2338  
631 aspGlyGlyValAlaValAspAlaGlyValAspThrGlnTrpGlyAlaL 648  
2339 TTAGTGCAGCGCA.....GACACGCGACTATACGGTTTACG 2373  
648 euMetAlaAspSerSerGlyGlnHisGlnAspGluGlySerThrLeuThr 664  
2374 CGCAAGCGCCACCAAAACGCGCAACCTCAGCCCTCGTGGGCAATGCCCAAGC 2423  
665 LysThrGlyAlaGlyThrLeuGluLeuThrAlaSerGlyThrThrGlnSe 681  
2424 AACA.....TTAATCAAGCCACATTAAACGGCACACATCG..... 2460  
681 rAlaValArgValGluGluGlyThrLeuLysGlySerValAlaAspIleL 698

2461 .....GCTTCGCACAATGCTCTATTATTAATCTA 2487  
698 euProTyAlaSerSerLeuTrpValGlyAspGlyAlaThrPheValThr 714  
2488 .....AGCAACAACGGCGTACAAACGGCAGTCT 2516  
715 GlyAlaAspGlnAspIleGlnSerIleAspAlaIleSerSerGlyThrIle 731  
2517 GACGCTTTCGCACAACGCTAAGCAACAACGTAAAGCCATTCCGCACTCAACG 2566  
731 eAspIleSerAspGlyThrValLeuArgLeuThr..... 742  
2567 GCAATGTCTCCTAGCGGATAAGGCAGTATTCCATTTTGAAAAACACGCCG 2616  
742 ..... 742  
2617 TTTACCGGAANAATCAGCGCGCAGGATACGGCAATTACACTTAAAGA 2666  
743 .....GlyGlnAspThrSerValAlaLeuAsnAl 752  
2667 CAGC.....GAATGGAGCTGCGCTCGGCACGGAA..... 2697  
752 aSerLeuPheAsnGlyAspGlyThrLeuValAsnAlaThrAspGlyValT 769  
2698 .....TTAGGCAATTAAAC..... 2712  
769 hrLeuThrGlyGluLeuAsnThrAsnLeuGluThrAspSerLeuThrTyr 785  
2713 CTTGACACGCCCAATTACACTCAATTCGCGCTATCGACAGATCGCGC 2762  
786 LeuSerAsnValThrValAsnGlyAsnLeuThrAsnThrSerGlyAlaVa 802  
2763 AGGCGCGCAACACCGCAGTCCGCGAGATCGCGCGCGCGCGCTCCGCGC 2812  
802 lserLeuGlnAsnGlyValAlaGlyAsp..... 811  
2813 GTTCCTATTATTCGTTACGCCGCCAATCTGGCAGCAATCCCGTTTCAAC 2862  
811 ..... 811  
2863 AGCGCTGACGGTAAACGGCAATTTGAACGGTCAGGACATTCGCTTTAT 2912  
812 ThrLeuThrValAsnGlyAspTyrThrGlyGlyThrLeuLeuLeuAs 828  
2913 GTCGGAATCTTCGCGC.....TACCGCAGCGGCAAAATTCAAAGCTGGCGG 2956  
828 pserGluLeuAsnGlyAspAspSerValSerAspGlnLeuValMetAsnG 845  
2957 AAAGTTTCGAAGGCACTTACACCTTGGCTGTCAACAAT...ACCGCG... 3000  
845 lYAsnThrAlaGlyAsnThrThrValValAsnSerIleThrGlyIle 861  
3001 AACGAACCCGTAAGTCTCGACGAATTGACGGTAGTGAAGGAAGAACAA 3050  
862 GlyGluProThrSerThrGlyIleLysValValAspPheAlaAlaAspPr 878  
3051 CACACGGCTGTCGGAATCTTAATTTACCCCTCCAAAACGAA...CACG 3097  
878 othrGlnPheGlnAsnAsnAlaGlnPheSerLeuAlaGlySerGlyTyrV 895  
3098 TCGATGCGCGCATGGCGTTATCAGCTTATCCGCAAGACGGCGAGTTC 3147  
895 alaAsnMetGlyAlaTyrAspTyrThrLeuValGluAspAsnAsnAspTrp 911  
3148 CGCCTGCATATTCGCGTCAAGCAACAGAGCTTTCCGCACAACTCGCAA 3197  
912 TyrLeu..... 913  
3198 GCGCGGAGAAACAGAGCGCGCTTGACGGCAAAAACAGGCACAACTTCGCG 3247  
913 ..... 913

78 LaVal.....Phe 80

772 TTACCAACAGGAGGCTCATTT.....GCGACAGTGGCTC 806  
||| |||:|||||

81 LeuGlnGlyAlaGluPheThrLeuSerProGluAsnGluThrGlyLeu 97  
||| |||:|||||

807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGG 856  
:|||||

97 tThrLeuPheAlaAsnAsnThr.....ValSerGlyG 108  
:|||||

857 TATTGCAACAGGCAACCCCTATATAGAAAAGCAATGGGTCCACACTA 906  
:|||||

108 LuTyrrAsnAsnGlyGlyAlaIlePheAlaTyrrGluAsnSerThrLeuAsn 124  
||| |||:|||||

907 GTTCGTAAGAATTGGTTCATGATGAATCTTCTGCTGAGATACCCATT 956  
:|| |||:|||||

125 LeuThr.....AspValIlePheSerGlyAsnValAlaG 136  
:|||||

957 AGTATTCTACGAACACATCAAAATGGCAATCTTTTAAACGACATA 1006  
:|||||

136 yGlyTyrr.....GlyGlyAlaIleTyrrSerSerGlyT 147  
:|||||

1007 ATAATGGCGCAGGAAAAATCGATGCCAACATAAACACTATTCTACCT 1056  
:|||| |||:|||||

147 hrAsnAspThrGlyAlaIleAsp.....154

1057 TATGAGTTAAAAACAGCAACCGTTCATGTTTAATGTTTCTTTATCCGA 1106  
|||:|||||

155 .....LeuArgValThrAsnAlaValPheArgAsnAsnIleAlaAsnAs 169  
:|||||

1107 GACAGCAAGAGAACCTGTTTATCATGTCGAGGTGGGTCAACAGTTATC 1156  
:|||||

169 pGlyGlyGlyAlaIleIleThrIleAsnAsnAspIleTyrrLeuSerA 186  
:|||||

1157 GACCAGACTGAATAATGAGAA.....AATATTCCTTTATGTAC 1197  
:|||||

186 spAspValPheAsnAsnAsnGlnAlaTyrrThrSerThrSerTyrrSerAsp 202  
:|||||

1198 AAGGAAAGGTGAATTCATCTATTACAGCAAC.....1230  
||| :|||:|||||

203 GlyAspGlyGlyAlaIleAspValThrAspAsnAsnSerAspSerIlyshi 219  
:|||||

1231 .....ATCAACCAAGCGCGCGGCTTTGTATTG 1261  
:|||||

219 sProSerGlyTyrrThrIleIleAsnThrAla.....230

1262 AGGGTAATTTTACGGTCTCCCTAAAAACACGAAACGTCGCAAGCGCG 1311  
:|||||

231 .....PheThr.....AsnAsnThrAlaGluGlyTyrr 239  
:|||||

1312 GCGTTCATATC....AGTGATGCGACTACGTTACTTGGAAAGTAAACGG 1358  
||| |||:|||||

240 GlyGlyAlaIleTyrrThrAsnSerAlaThrAlaProTyrrLeuIleAspI 256  
:|||||

1359 CGTGCAACACGCGCTGTCCAAAATCGCAAAAGCAGCGTGTGGTTC 1408  
:|||||

256 eSerValAspSerTyrrSerGlnAsnGly...GlyValLeuValAspG 272  
:|||||

1409 AA.....GCCAAGGGGAAAAACCAAGGCTCGGTGACG.....1440  
:|| ||| :|||:|||||

272 LuAsnAsnSerAlaAlaGlyTyrrGlyAspGlyProSerSerAlaAlaGly 288  
:|||||

1441 .....GTGGGCGACGG 1451  
:|||||

289 GlyPheMetTyrrLeuGlyLeuSerGluValThrPheAspIleAlaAspG 305  
:|||||

1452 TAAAGTCATCTTAGATCAGCAGCGGACGATCAAGCGCAAAACAAAGCCT 1501  
|||:|||||

305 yLysThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerI 322  
:|||||

1502 TTATGTAATCGCTTGGTC.....AGCGGACGGGACGGTGCACACTG 1545  
:|||||

322 leAlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValIleu 338  
:|||||

```

537 AACCGGGAAGTTCATATCATATATTGCAAGCGCATATCT..... 675
      ::::|||||
46 SerArgGlnSerLeuSerGlyLeuThrGlnValTrpSerIleAlaSpG1 62
      ::::|||||
676 ...TGGTCGTGCTGGTCATACCTTTGCACAAAATGCATCAGTCTG 721
      :|||||||
62 yGlnTrpLeuValPheSerAsp...MetThrAsnAsnAlaSerGlyGlyA 78
      :|||||||
722 GCACAGTCAACTTAGGTAGCGAAATAATTAAACATAGCCCATATGTT 771
      ::::
      |||

```



1546 AATCGCGATAATACAGTTCACACCCGACAACTCTATTTCGGCTTCGGCG 1595  
 339 AsnAlaAspAsn..... 342  
 1596 CGGACGTTTGGATTTCACGGGCGATTCGCTTCGTTCCACCGCATTCAAA 1645  
 343 .....A 343  
 1646 ATACCGATGAAGGGCGGATGATTGTCAACCAATCAAGACAAAGAAATCC 1695  
 343 snasphethrGlyGluMetGlnIleGluAsnGlyGlu..... 355  
 1696 ACCGTTACATACAGGCAATAAGATATTACTACACCGGCAATACAA 1745  
 356 ...ValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHi 371  
 1746 CAACITGGATACAAAAGAAATTCGCTTACACCGTTGGTTGGCGAGA 1795  
 371 sCysGlnAspAspProGlnAsp...CystyrGlyLeuThrIleGlySerI 387  
 1796 AAGATCAACCAACGAACGGCGGCTCAATCTGAATACCAACCGGAA 1845  
 387 leasplysyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln 403  
 1846 GAAGCGGATCGCACTTA.....CTGCTTCGGCGGCAACAATTTAA 1889  
 404 ThrPheAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAs 420  
 1890 C.....GCAATATCACCAACAAACGGCAAACTGTTTTCAGCGGCA 1933  
 420 pAlaGlyGlyAsnValThrValAsnGlnGlySer.....PheAlaGly. 434  
 1934 GACCGACCGCGACGCTCAATCAATTTAGGAAGCGGTTGGTCAAAAATG 1983  
 435 .....ThrIle 436  
 1984 GAGGTATCCCAAGAGAAATCGTGGGACACGATTTGGATCGACCG 2033  
 437 GluGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVal..... 451  
 2034 CACATTTAAAGCGAAATCTTCATATTCAGGCGGACAAAGCGGTGTTT 2083  
 452 LeuAlaGlyAlaGlnSerMetAlaLeuThrGlyAspIleValValAspA 468  
 2084 CCGCAATGTTGCCAAGTGGAGGCGAT.....TGGCATTTA 2121  
 468 laGlyAlaValLeuSerLeuGluGlyAspAlaAlaAspLeuAlaLeu 484  
 2122 AGCAATCACGCCCAAGCAGTTTCGGTGTGCGCACCGCATCAAAAGCCACAC 2171  
 485 GlnAspAspProGlnSerIle.....ValLeuAsnGlyGlyMetLeuAs 499  
 2172 AATCTGTACAGTTTCGAGCTGGAGCGGTCTGACAAAGTTGT.....A 2212  
 499 pLeuSerAspPheSerThrTrpGlnSerGlyThrSerTyrLysAspGlyL 516  
 2213 CCGAAAAACCAATACCGACGATAAGTGTGCTTCATTCGACCAAGACC 2262  
 516 euGluValSerGlySerSerGlyThrValIleGlySerGlnAspValVal 532  
 2263 GACATCAGA..... 2271  
 533 AspLeuAlaGlyGlyAsnAspMetHisIleGlyGlyAspGlyLysAspGI 549  
 2272 .....GGCAATGTGAGCTTGGCC...G 2290  
 549 yValTyrValValIleAspAlaGlyAspGlyGlnValSerLeuAlaAsnA 566  
 2291 ATCAGGCTCATTTAAATCTACAGGACTTGCACACTCAACGGCAATCTT 2340  
 566 spAsnGlnTyrLeuGlyThrThrGlnIleAla.....SerGlyThrLeu 580

2341 .....AGTGCAGCGGAGACACGCACTAT..... 2364  
 581 MetValSerAspAsnSerGlnLeuGlyTyrThrHisTyrAsnArgGlnVa 597  
 2365 .....A 2365  
 597 lilePheThrAspLysProGlnGluSerValMetGluIleThrAlaAsnV 614  
 2366 CGGTTAGCGGCAAGCCACC..... 2385  
 614 alaSpThrArgSerThrThrGluHisGlyArgAspIleGluMetArg 630  
 2386 CAACACCGCAACCTCAGCCTCGTGGCAATGCCCAAGCAACATTTAATCA 2435  
 631 AlaaspGlyGluValAlaValAspAlaGlyValAspThrGlnTrpGlyAl 647  
 2436 AGCCACATTAACGGCAACATCGGCTTCGGACATGCTTCATTTAATC 2485  
 647 aLeuMetAlaAspSerSerGlyGlnHisGlnAspGluGlySer...ThrL 663  
 2486 TAACCAACACCGCGCTACAAACGGCAGTCTGACGCTTCGCACACCGCT 2535  
 663 euThrLysThrGlyAlaGlyThrLeuGluLeuThrAlaSerGlyThrThr 679  
 2536 AAGCA.....AACGTAAGCCATTCCGCACTCAACGCAATGTCCTCCT 2579  
 680 GlnSerAlaValArgValGluGluGlyThrLeuGlnGlyAspVal..... 694  
 2580 AGCGATTAAGCAGTATTCATTTGAAACAGCGCTTTTACCGGAAAA. 2628  
 695 AlaAsp.....IlePheProTyrAlaSerSerLeuTrpValGlyAspG 709  
 2629 .....ATCAGCGGC..... 2637  
 709 lyAlaThrPheValThrGlyAlaAspGlnAspIleGlnSerIleAspAla 725  
 2637 ..... 2637  
 726 ThrSerSerGlyThrIleAspIleSerAspGlyThrValLeuArgLeuTh 742  
 2638 .GGCAAGGATACGGCATTACACTTAAAGACAGC.....GAAT 2674  
 742 rGlyGlnAspThrSerValAlaLeuAsnAlaSerLeuPheAsnCysaspG 759  
 2675 GGACGCTCGCTCGGCGACGGAA.....TTAGGCAATTTAAAC 2712  
 759 lyThrLeuValAsnAlaThrAspGlyValThrLeuThrGlyGluLeuAsn 775  
 2713 .....CTTGACACGCCACCATTAC 2732  
 776 ThrAsnLeuGluThrAspSerLeuThrTyrLeuSerAsnValThrValAs 792  
 2733 ACTCAATTCGCGCTATCGACACGATGCGGCGCGCAACCGCGAGTG 2782  
 792 nGlyAsnLeuThrAsnThrSerGlyAlaValSerLeuGlnAsnGlyValA 809  
 2783 CGGCAAGTGGCGCGCGCGCGCTTCGCCCTTATTCCTGTTACG 2832  
 809 laGlyAsp..... 811  
 2833 CGGCCAACTTCGCGAGATCCCGTTTCAACACGCTGACGCTAAACGGCAA 2882  
 812 .....ThrLeuThrValAsnGlyAs 818  
 2883 ATTGAACGCTCAGGAACATTCGCTTTATGTCGGAACCTCTTCGGC.... 2928  
 818 pTyrThrGlyGlyThrLeuLeuLeuAspSerGluLeuAsnGlyAspA 835  
 2929 ..TACCGCACCGCAAAATTAAGCTGGCGGAAAGTTCCGAGGACCTTAC 2976  
 835 spSerValSerAspGlnLeuValMetAsnGlyAsnThrAlaGlyAsnThr 851  
 2977 ACCTTGGCTGTCAACAAT...ACCGGC....AACGAACCGGTAAAGTCTCCA 3020



A:Residues: 1-1250 <STO>

A:Cross-references: GB:AE005174; NID:g12516568; PIDN:AA057362.1; GSPDB:GN00145; UWGP:Z34

A:Experimental source: strain 0157:H7, substrain EDL933

C:Genetics:

A:Gene: yfaL

alignment\_scores:

Quality: 290.50 Length: 1484  
Ratio: 0.441 Gaps: 73  
Percent Similarity: 44.340 Percent Identity: 19.946

alignment\_block:

US-09-303-518D-653 x F95862 ..

Align seg 1/1 to: F85862 from: 1 to: 1250

```

637 AACCGCGAAAGTTCATATCATATTCAGAGCGCATATCT..... 675
|||||..... |||.....
46 SerArgGlnSerLeuSerGlyIleThrGlnValTrpSerIleAlaAspG1 62
676 ....TGCGTCGCGGTGGCATACCTTTGCACAAATGGATCGAGTGTG 721
|||||..... |||.....
62 yGlnTrpLeuValPheSerAsp...MetThrAsnAsnAlaSerGlyGlyA 78
722 GCACAGTCAACTAGGTAGCGGAAAAAATAAACATAGCCCATATGTTT 771
|||||..... |||.....
78 IVal.....Phe 80
772 TTACCAACAGAGGCTCATTT.....GGCAGAGTGGCTC 806
|||||..... |||.....
81 LeuGlnGlnGlyAlaGluPheThrLeuSerProGluAsnGluThrGlyMe 97
807 ACCAATGTTTATCTATGATGCCCAACAAAGTGGTTAATTAATGGG 856
|||||..... |||.....
97 tThrLeuPheAlaAsnAsnThr.....ValSerGlyG 108
857 TATTGCAACAGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTA 906
|||||..... |||.....
108 LuTrpAsnAsnGlyGlyAlaIlePheAlaLysGluAsnSerThrLeuAsn 124
907 GTTCGTAAGATGGTCTCTATGATGAATCTTTGTGGAGATACCCATTC 956
|||||..... |||.....
125 LeuThr.....AspValIlePheSerGlyAsnValAlaG1 136
957 AGTATCTACGACCAACATCAAAATGGGAATACATTTTAAACGCAATA 1006
|||||..... |||.....
136 yGlyTyr.....GlyGlyAlaIleTyrSerSerGlyT 147
1007 ATAATGGCGGAGAAAAATCGATGCCAAACATAAACACTATCTCTACCT 1056
|||||..... |||.....
147 hrAsnAspThrGlyAlaIleAsp..... 154
1057 TATAGATATAAACACAGACCGTTCATTTGTTAATGTTCTTTATCCGA 1106
|||||..... |||.....
155 .....LeuArgValThrAsnAlaValPheArgAsnAsnIleAlaAsnAs 169
1107 GACAGCAGACAGACCTGTTTATCATCTGCAGTGGGGTCAACAGTTATC 1156
|||||..... |||.....
169 pGlyLysGlyAlaIleTyrThrIleAsnAsnAspIleTyrLeuSerA 186
1157 GACCCAGACTGAATAATGGAGAA.....AATATTTCTTTTATTGAC 1197
|||||..... |||.....
186 spAspValPheAsnAsnAsnGlnAlaTyrThrSerThrSerTyrSerAsp 202
1198 AAAGGAAAGGTGAATGATCTTACACGAAC..... 1230
203 GlyAspGlyGlyAlaIleAspValThrAspAsnAsnSerAspSerLysH1 219
1231 .....ATCAACCAAGCGCGGGGGTTCATTTGTTTATTTG 1261
|||||..... |||.....
219 sProSerGlyTyrThrIleIleAsnAsnThrAla..... 230

```

```

1262 AGGGTAATTTTACGCTCGCCTAAAAACACAGCAACGTGGCAAGCGCG 1311
|||||..... |||.....
231 .....PheThr.....AsnAsnThrAlaGluGlyTyr 239
1312 GCGGTTTCATATC...ACTGATGCGCAGTACGTTACTTGTGAAAGTAACGG 1358
|||||..... |||.....
240 GlyGlyAlaIleTyrThrAsnSerAlaThrAlaProTyrLeuIleAsp1 256
1359 CGTGGCAACAGCGCGCTGTCCAAATCGCAAGGCGCTCGTCTGGTTC 1408
|||||..... |||.....
256 eSerValAspAspSerTyrSerGlnAsnGly...GlyValLeuValAsp6 272
1409 AA.....GCCAAAGGGGAAACCAAGGCTCGTCTGACG..... 1440
|||||..... |||.....
272 LuAsnAsnSerAlaAlaGlyTyrGlyAspGlyProSerSerAlaAlaGly 288
1441 .....GTGGCGCAGCG 1451
289 GlyPheMetTyrLeuGlyLeuSerGluValThrPheAspIleAlaAspG1 305
1452 TAAAGTCATCTTAGATCAGCAGCGGACGATCAAGCAAAAAACAAGCCT 1501
|||||..... |||.....
305 yLysThrLeuValIleGlyAsnThrGluAsnAspGlyAlaValAspSerI 322
1502 TTAGTGAATCGGCTTGGTC.....AGCGCAGGGGCGCGTGCACACTG 1545
|||||..... |||.....
322 LeAlaGlyThrGlyLeuIleThrLysThrGlySerGlyAspLeuValLeu 338
1546 AATCCGCAATATCAGTTCAACCCGACAACTCTATTTCGGCTTCGCGG 1595
|||||..... |||.....
339 AsnAlaAspAsn..... 342
1596 CGGACGTTTGGATTGAACGGCATTCGCTTCGTTCCACCGCATTCAAA 1645
343 .....A 343
1646 ATACCGATGAAGGCGCATGTTCTCAACACATCAAGCAAGATCC 1695
|||||..... |||.....
343 snAspPheThrGlyGluMetGlnIleGluAsnGlyGlu..... 355
1696 ACCGTTACCATTCACAGCAATAAAGATATTACTACAACCGGCAATAACA 1745
|||||..... |||.....
356 ...ValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrH1 371
1746 CAACTTGGATAGCAAAAAAGAAATTGCCTACACGGTGTGGTTGGCGAGA 1795
|||||..... |||.....
371 sCysGlnAspAspProGlnAsp...CysTyrGlyLeuThrIleGlySerI 387
1796 AAGATGCAACCAACGACGCGGCTCAATCTGAATTACCAACCGGAA 1845
|||||..... |||.....
387 LeAspLysTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln 403
1846 GAAGCGGATCGCACTTTA.....CTGCTTTCCGCGGGAACAAATTTAAA 1889
|||||..... |||.....
404 ThrPheAlaHisSerLeuThrGlyPheGlnAsnGlyThrLeuAsnIleAs 420
1890 C.....GGCAATATTCAGCAAAACAAACGCAAACTGTTTTCACGGCA 1933
|||||..... |||.....
420 pAlaGlyGlyAsnValThrValAsnGlnGlySer.....PheAlaGly. 434
1934 GACCCAGACCGCACCCCTACATCATTTAGNAGCGGTGGTCAAAAATG 1983
435 .....ThrIle 436
1984 GAAGTATCCCAAGAGAGAATCGTGTGGGACACGATTTGGATGACCG 2033
|||||..... |||.....
437 GluGlyAlaGlyGlnLeuThrIleAlaGlnAsnGlySerTyrVal..... 451
2034 CACATTTAAAGCGGAAAACTTCATATTCAGGCGGACAAAGCGGTGGTT 2083
|||||..... |||.....
452 .LeuAlaGlyAlaGlnSerMetAlaLeuThrGlyAspIleValValAspA 468
2084 CCCGCAATGTTGCCAAAGTGGAGGCGAT.....TGGCATTTA 2121

```

```

::
468 laGlyAlaValLeuSerGluGlyAspAlaAlaAspLeuAlaLeu 484
:::
2122 AGCAATCACGCCCAAGCAGTTTCGGTGTGCGACCGCATCAAGCCACAC 2171
:::
485 GlnAspAspProGlnSerIle.....ValLeuAsnGlyMetLeuAs 499
:::
2172 AATCTGTACACCTTCGACGTGGAGGCTCTGCAAGTTGT.....A 2212
:::
499 pLeuSerAspPheSerThrThrGlnSerGlyThrSerTyrLysAspGlyL 516
2213 CCGAAAAAACATTACCGACGATAAAGTATTGCTTTCATTGACCAAGACC 2262
:::
516 euGluValSerGlySerSerGlyThrValIleGlySerGlnAspValVal 532
2263 GACATCAGA..... 2271
533 AspLeuAlaGlyGlyAsnAspMetHisIleGlyGlyAspGlyLysAspGI 549
2272 .....GGCAATGTACGCTTGGC...G 2290
549 yValTyrValValIleAspAlaGlyAspGlyGlnValSerLeuAlaAsnA 566
2291 ATCAGCGCTATTAAATCTCACAGGACTTGGCCACACTCAACGGCAATCTT 2340
:::
566 spAsnGlnTyrLeuGlyThrGlnIleAla.....SerGlyThrLeu 580
2341 .....AGTCGAGCGGAGACACGCATAT..... 2364
581 MetValSerAspAsnSerGlnLeuGlyTyrThrHisTyrAsnArgGlnVa 597
2365 .....A 2365
597 lIlePheThrAspLysProGlnGluSerValMetGluIleThrAlaAsnV 614
2366 CGGTTACGCCCAAGCCACC..... 2385
614 aIAspThrArgSerThrThrThrGluHisGlyArgAspIleGluMetArg 630
2386 CAAACGGCAACCTCAGCCTCGTGGCAATGCCCAAGCACATTTAAACA 2435
:::
631 AlaAspGlyValAlaValAlaAspAlaGlyValAspThrGlnTrpGlyAl 647
2436 AGCCATTAAACGGCACACATCGGCTTCGGACAATGCTTCAATTAAATC 2485
647 aLeuMetAlaAspSerSerGlyGlnHisGlnAspGluGlySer...ThrL 663
2486 TAAGCAACAACCGCTACAAAACGGCAGTCTGACGCTTTCGACACACGCT 2535
:::
663 euThrLysThrGlyAlaGlyThrLeuGluLeuThrAlaSerGlyThrThr 679
2536 AAGGCA.....AAGTAAGCCATTCCGCACTCAACGGCAATGCTCCCT 2579
:::
680 GlnSerAlaValArgValGluGluGlyThrLeuGlnGlyAspVal..... 694
2580 ACCGATTAAGCAGATTCCATTTTGAACACAGCGCTTTTACCGGAAAA. 2628
:::
695 .AlaAsp.....IlePheProTyrAlaSerSerLeuTrpValGlyAspG 709
2629 .....ATCAGCGGC..... 2637
709 lYAlaThrPheValThrGlyAlaAspGlnAspIleGlnSerIleAspAla 725
2637 ..... 2637
726 ThrSerSerGlyThrIleAspIleSerAspGlyThrValLeuArgLeuTh 742
2638 .GGCAAGATACGCATTACATTTAAAGACAGC.....GAAT 2674
:::
742 rGlyGlnAspThrSerValAlaLeuAsnAlaSerLeuPheAsnCysAspG 759
2675 GGAGCTCCCTCGGGCACGGAA.....TTAGCAATTAAAC 2712
:::

```

```
3568 AATAGCGGTTGAGTCAATTTTCGCCCGCTCAACAGCGTTTTCGCGT 3617
      ::::: ::::: |||||::: ||::
956 .....AlatValGlyTyrLeuAsnLeuArgAlaAl 968
      ::::: ::::: |||||::: ||::
3618 ACAGGACGAATGGACCGCGTGTGGCAAGACCGCGCAACGCGGTTT 3667
      ::::: ::::: |||||::: ||::
968 a.....AsnGlnAlaPheMetMetGluArgAspHisAlaG 981
      ::::: ::::: |||||::: ||::
3668 GGACAAGCGG.....ATCCGGGACACCAACAC 3696
      ::::: |||||
981 lyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyAspTyrHis 997
      ::::: |||||
3697 TACCGTCCGCAAGATTCGCGCTACCGCCCAACACCGACCTCGGCCA 3746
      |||||
998 TyrThrAlaAlaGlyGlnLeuAla.....GlnHisGluAspThrSerTh 1012
      |||||
3747 AATCGGTATGCAGAAAAACCTCGGACGCGGCGC..... 3780
      ::::: |||||
1012 rValGlnLeuSerGlyAspLeuPheSerGlyArgTrpGlyThrAspGlyG 1029
      ::::: |||||
3781 .....GTCGGCATCTG.....TTTCCGACAAACCGGACCGGAMAC 3816
      ::::: |||||
1029 lnrTrpMetLeuGlyIleValGlyTyrSerAspAsnGln...GlyAsp 1044
      ::::: |||||
3817 ACCTTCGACGCGCATCGCAACTCGGCA...CGGCTTCCCGACGGTGC 3863
      ::::: |||||
1045 SerArgSerMetThrGlyThrArgAlaAspAsnGlnAsnHisGly.. 1060
      ::::: |||||
3864 CGTTTTCGGCAATACGGCATCGGAGGTTTCGACATCGGCATCGCGCG 3913
      |||||
1061 .....TyrAlaValGlyLeuThrSerSerTrpPheGlnHisG 1073
      ::::: |||||
3914 GC.....CGGGT 3921
      |||
1073 lyLysGlnLysGlnGlyAlaTrpLeuAspAsnTrpLeuGlnTyrAlaTrp 1089
      |||
3922 TTTAGTAGC.....GGCAGCTTTCAGACGCGCATCAGAGGCAAAATCCG 3965
      |||||
1090 PheSerAsnAspValSerGluHisGluAspGlyValAsp..... 1102
      ::::: |||||
3966 CCGCGCGGTGCTGCATTAC.....GGCATTCAGCAAGATACCGCG 4006
      |||||
1103 .....HisTyrHisSerSerGlyIleIleAlaSerLeuGluA 1115
      ::::: |||||
4007 CAGTTTC.....GGCGGATTCCGGATCCGAACCGCACATC 4041
      |||||
1115 lagGlyTyrGlnTrpLeuProGlyArgGlyValValIleGluProGlnAla 1131
      ::::: |||||
4042 GCGCAGCGCGTATTTCGTCCAAAAGCGGATTACCGATACGAAAC.. 4089
      ::::: |||||
1132 GlnValIleTyrGlnGlyValGlnGlnAspPheThrAlaAlaAsnAr 1148
      ::::: |||||
4090 .....GTCAATATCGCACCCCGCGGCTTG 4114
      ::::: |||||
1148 gAlaArgValSerGlnSerGlnGlyAspAspIleGlnThr..... 1161
      ::::: |||||
4115 CATTCAACCGGTACCGCGGGCATTAAGCCAGATTATTCAATCAACCG 4164
      |||||
1162 .....ArgLeuGlyLeuHisSerGluTrp.....ArgThr 1171
      ::::: |||||
4165 GCGCAACACATTTCCATCGCCTTATTTCAGCTGTCC...TATACCGA 4211
      |||||
1172 AlaValHis.....ValIleProThrLeuAspLeuAsnTyrThrHisAs 1186
      ::::: |||||
4212 TGCGCTTCGGCAAGTCCGACGCGGTCAATACCGCGGTATTGGCGC 4261
      |||||
1186 pProHisSerThrGluIleGluGluAspAlaSerThr.....IleSerA 1201
      ::::: |||||
4262 AGGATTTTCGCAAAACCGC...AGTCGGAATGGCGGTAAACGCCGA 4308
      ::::: |||||
1201 spAspAlaValLysGlnArgGlyGluIleLysValGlyValThrGlyAsn 1217
      ::::: |||||
```

```
4309 ATCAAAGGTTTCACGCTGTCCCTCCAGCTCGCGCGCC.....AAGG 4352
      |||||
1218 IleSerGln...ArgValSerLeuArgGlySerValAlaTrpGlnLysG 1233
      ::::: |||||
4353 GCCCAATTTGAGCGGACGCGCACGCGGCGGATCAAAATTAGGCTACCGCT 4402
      |||||
1233 ySerAspAspPheAlaGlnThrAlaGlyPheLeuSerMetThrValLys 1250
      ::::: |||||
4403 GG 4404
      ||
1250 rp 1250

seq_name: pir2:B43855

seq documentation_block:
high-molecular-weight surface-exposed protein - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C:Accession: B43855
R:Barenkamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A:title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable
detella pertussis.
A:Reference number: A43855; MUID:92192797
A:Accession: B43855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1477 <BAR>
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:89237, NCBI:89240)
```

```
alignment_scores:
Quality: 279.50 Length: 1319
Ratio: 0.437 Gaps: 61
Percent Similarity: 48.446 Percent Identity: 20.015

alignment_block:
US-09-303-518D-653 x B43855 ..
Align seg 1/1 to: B43855 from: 1 to: 1477
```

```
706 AATGATCAGTGTGGCGACAGTCAACTAGTAGCGGAAAAAATAACA 755
      |||||
165 AsnGlyPheThrAlaSerThrLeuAspIleSerAsnGluAsnIleLys.. 180
      ::::: |||||
756 TAGCCCATATGTTTTTACCACAGGAGGCTCATTTGGCGACAGTGCT 805
      |||||
181 .....A 181
      ::::: |||||
806 CACCAATGTTTATCTATGATGCCCAAAAGTGGTTAATTAATGGG 855
      ::::: |||||
181 laArgAspPheThrPheGluGlnThrLysAspLysAlaLeuAlaGlu... 196
      ::::: |||||
856 GTATTGCAACAGGACCCCTATATAGAAAAGCAATGGCTTCCAGCT 905
      ::::: |||||
197 IleValAsnHisGlyLeuIleThrValGlyLysAspGlySerValAsnLe 213
      ::::: |||||
906 AGTTGCTAAGATGTTCTTATGATCAATCTTCTGGAGATACCCATT 955
      |||||
213 uile..... 214
      ::::: |||||
956 CAGTATTCTACGACACATCAAAATGGGAAATACCTTTTAAACGACAAT 1005
      ::::: |||||
215 .....GlyGlyLysVal..... 218
      ::::: |||||
1006 AATAATGGCGAGGAAATATCGATGCCAATAAACACTATTCTCTACC 1055
      ::::: |||||
219 ...LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLe 234
      ::::: |||||
1056 TTATAGATTAAACACAGAACCGCTTCAATGTTTAAATGTTTCTTATCG 1105
      |||||
234 uAlaGlyGlnLysIleThrIleSerAspIleIleAsnProThrIleThrT 251
      ::::: |||||
```

1106 AGACAGCAAGAGAACCT.....GTTTAT 1128  
111 ySerrileAlaAlaProGluAsnGluAlaValAsnLeuGlyAspIlePhe 267  
1129 CATGCTCCAGCTGGGTCAACAGCTTATCCAGCCAGACTGAATAATGAGA 1178  
268 AlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleArgAsnGlnG 284  
1179 AAATATTTC.....TTTATTGACAAAGGAAAA...GGTGAATGATAC 1219  
284 yLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleVal 301  
1220 TTACAGCAACATCAACCAAGGC...GCGGGCGGTTTGTATTGAGGGT 1266  
301 euSerAlaLysGluGlyGluAlaGluIleGlyVal..... 313  
1267 AATTTTACGCTCGCTCAAAACACGAAACGTGGCAAGCGCGGGCT 1316  
314 .....IleSerAlaGlnAsnGlnAla...LysGlyGlyLysLe 326  
1317 TCATATCAGTGAGTGAGTACCGCTTACTTGGAAAGTAAACGGCGTGGCAA 1366  
326 uMetIleThr...GlyAspLysValThrLeuLys...ThrGlyAlaVal 341  
1367 ACCAGCCCTGTCCAAATCGCAAGGACGCGTGTG...GTTCAAGCC 1413  
341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357  
1414 AAGGGGAAACCAAGCGTCGTC.....AGCGTGGG 1445  
358 ArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLysLysThrSerLeuG 374  
1446 CGAGGTAAAGTATCTTAGATCAGCAGCGGACGATCAAGCAAAAC 1495  
374 uLysGlySerThrIleAsnValSerGlyLysGluLysGlyGlyAla 391  
1496 AAGCCTTTAGTGAATCGCTTGGTCAGCGGAGGCGGAGGTGCACTG 1545  
391 leValIleGlyAspIleAlaIleAspGlyAsnIleAsnAlaGlnGly 407  
1546 AATCGCGATAATCAGTTCAACCCGACAACTCTATTTCGGCTTCGCGG 1595  
408 SerGlyAsp.....IleAlaLysThrG 415  
1596 CGAGCTTTGGATTGACAGCGCATTCGTTTCGTC.....CACCGCA 1639  
415 yGlyPheValGluThrSerGlyHisThrLeuSerIleGluSerAsnAla 432  
1640 TTCAAAATACCGATCAAGGGCGATGTGTCAACCAACATCAAGACAA 1689  
432 leValLysThrLysGluThrLeu.....Asp 441  
1690 GAATCCACCGTTTACCATT.....ACAGG 1712  
442 ProAspAspValThrIleGluAlaGluAspProLeuArgAsnAsnThrG 458  
1713 CAATAAGATATTACTACACGGGAATACACAACTTGGATACGAAA 1762  
458 yIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLys 475  
1763 AA.....GAATT 1770  
475 yAsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnThr 491  
1771 GCCTACACAGGTGGTTTGGCGAAGAGATGCAACCAAA.....AC 1811  
492 LeuLysAsnAlaTyrThrMetAsnIleThrAlaSerArgLysLeuThrVa 508  
1812 GAACGGCGGCTCAATCTGAATTTACCAACCGGAAGACCGGATCCGACT 1861  
508 IAsnSerIleAsnIleGly.....SerAsnSerHisL 520

1862 TACTGCTT.....TCCGGCGGACAAATTTAAACGGC 1893  
520 euIleLeuHisSerLysGlyGlnArgGlyGlyValGlnIleAspGly 536  
1894 AATATCAGCAAAACAAACGCAAACTG.....TTTTTCAGCGCAGAC 1937  
537 AspIleThrSerLysGlyGlyAsnLeuThrIleThrSerGlyGlyTrpVa 553  
1938 GACACCGCAGCCCTACATCATTTAGGAAGCGGTGTCAAAATG.... 1983  
553 lAspValHisLysAsnIleThrLeuAspGlnGlyPheLeuAsnIleThr 570  
1984 .....GAAGCT..... 1989  
570 laAlaSerValAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaAla 586  
1990 .....ATCCACAGAGGAAATCGTGTGGACACACGATTTGGAT 2027  
587 AsnAlaLysIleValAlaGlnGlyThrValThrIleThrGlyGluGlyLy 603  
2028 CGACCGCACATTTAAAGCGGAAACCTTCCATATTTCAGGCG.... 2067  
603 sAsp.....PheArgAlaAsnAsnValSerLeuAsnGlyThrGlyLysG 618  
2067 ..... 2067  
618 lyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerGly 634  
2068 .....GGACAGCGGTGGTTTCCCGCAATTTGCCAAAGT 2102  
635 ThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLysAs 651  
2103 GGAAGCGGATTTGATTAAGCAATCATCGCCCAAGCAGTTTCGTGCTCG 2152  
651 nThrSerTyrTrpGlnThrSer..... 658  
2153 CACCGCATCAAGCCACACAACTGTGTACAGTTTCGGACTGG.....ACG 2196  
659 .....HisAspSerHis.....TrpAsnValSer 666  
2197 GGCTGACAAAGTTGTACCGAAAAACCATTTACCGACGATAAAGTATTCG 2246  
667 AlaLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSe 683  
2247 TTCAATTGACGAAAG.....ACCGACATCAGAGCAATGTCAGCCTTG 2287  
683 rSerAsnSerLysGlyLeuThrThrGlnThrArgSerSer..... 696  
2288 CCGATCAGCTCATTTAAATCTCACAGGACTTCCACACTCAACGCAAT 2337  
697 .....AlaGlyValAsnPheAsnGly.....ValAsnGlyAsn 707  
2338 CTTAGT.....GCAGCGGAGACACGCACTATACGTTTACGCG 2375  
708 MetSerPheAsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysPr 724  
2376 CAACGCCACCCAAAAC.....GCCA 2395  
724 oAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 741  
2396 ACCTCAGCCTCTGGGCAATGGCCAAAGCAACATTTAAT...CAAGCCACA 2442  
741 snIleThrAlaThrGlyGlySerValPhePheAspIleThrAlaAsn 757  
2443 TTAACGGCAACACATCGCTTCGGACAATGCTTCATTTTAACTAAGCAA 2492  
758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774  
2493 CAACGGCTACAAACGGCAGTCTGACGCTTCCGACACACGCTAAGGCA. 2541  
774 nGlyAla.....AsnPheThrLeuAsnSerHisValArgGlyA 787  
2542 .....ACGTAAGCCATTTCGCACTCAACGGC 2568



hemolysin A precursor - Proteus mirabilis

C:Species: Proteus mirabilis

C:Date: 03-Aug-1990 #sequence\_revision 03-Aug-1990 #text\_change 08-Oct-1999

C:Accession: A35140

R:Uphoff, T.S.; Welch, R.A.

J. Bacteriol. 172, 1206-1216, 1990

A:Title: Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin gene

A:Reference number: A35140; M01D:90170827

A:Accession: A35140

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1577 <UPH>

A:Cross-references: GB:M30186; NID:g150888; PIDN:AAA25657.1; PID:g150890

alignment\_scores:

Quality:	277.50	Length:	1715
Ratio:	0.373	Gaps:	88
Percent Similarity:	43.382	Percent Identity:	19.767

alignment\_block:

US-09-303-518D-653 x A35140 ..

Align seg 1/1 to: A35140 from: 1 to: 1577

```

40 AAAGCCCTTAAACCGCGCATCCGCTTCTCGCCCGCTTACTTAGCCAT 89
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
7  LysLeuSerProSerGlyArgLeuAlaAlaSerLeuAlaIlePheVa 23
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
90 ATGCGCTGCTGCTC.....GGCATCTGCCCAAGCCCGGGGGAG 130
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
23 IserLeuAsnAlaTyrGlyAsnGlyIleValProAsp.....AlaGlyH 38
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
131 AC.....TACAAACAAAGGGAGTGTGTC 132
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
38 IserGlnGlyProAspValSerAlaValAsnGlyGlyThrGlnValIleAsn 54
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
133 .....ACTATTTCGGCATCACTACCAATCTATCGCGACTT 170
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
55 IleValThrProAsnAsnGluGlyIleSerHisAsnGlnTyrGlnAspPh 71
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
171 TGCCCAAAATAAGGCAAG...TTTGCAGTCGGCGGAAAGATATTGAGG 217
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
71 e.....AsnValGlyLysProGlyAlaValPheAsnAsnAlaLeuGluA 86
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
218 TT.....TACAAACAAAGGGAGTGTGTC 243
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
86 IaGlyGlnSerGlnLeuAlaGlyHisLeuAsnAlaAsnSerAsnLeuAsn 102
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
244 GGCAATCGATGACGAAAGCCCGATGATTGTTTCTGTGGTATCGCG 293
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
103 GlyGlnAla.....AlaSerLeuIleLeuAsnGluValValSerAr 116
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
294 TAAC.....GGCGTGGCGGCAT 310
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
116 GAsnProSerPheLeuLeuGlyGlnGlnGluValPheGlyIleAlaAla. 132
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
311 TGGCGGGCGATCAATATATTGTGCGCGTGGCAGACATAAC..... 348
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
133 .....GluTyrValLeuSerAsnProAsnGlyIleThrCysAsp 145
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
349 .....GGCGGCTATACAATGTTGATTTTGTGGGAGGGAACAATCC 392
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
146 GlyCysGlyPheIleAsnThrSerArgSerLeuValValGlyAsnPr 162
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
393 C.....GATCAGCAGCCGCTTTCTTACCAAAATTGTGAAGAATA 433
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
162 oPhePheGluAsnGlyGlnLeuLysGlyTyrSerThrLeuAsnAsnThrA 179
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
434 AT...TATAAGACGGGACTAACGCCCATCTTATGGCGGC.....GAT 474
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
179 snLeuLeuSerLeuGlyLysAsnGlyLeuAsnThrThrGlyLeuLeuAsp 195
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

475 TATCATATGCGCGCTTTGCAAAA.....TTTGTCTACAGATCGAGAACC 518
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196 LeuIleAlaProArgIleAspSerArgGlyLysIleThrAlaAlaGluIle 212
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
519 TGTGTGACATGACC..... 531
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
212 eSerAlaPheThrGlyGlnAsnThrPheSerGlnHisPheAspIleLeuS 229
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
532 .....AGTTATATGATGGGTGGAAATACGCTGATTATA 564
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
229 erSerGlnLysProValSerAlaLeuAspSerTyrPhePheGlySerMet 245
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
565 AATAAATACCTGATCGTGTTCGAATC.....GGACGACAG 599
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
246 GluSer.....GlyArgIleArgIleIleAsnThrAlaGluGlySerGl 260
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
600 CAGACAATAT.....TGGCGGTCTGATGAAGAC.....GAAC 631
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
260 yValLysLeuAlaGlyLysPheThrAlaAspAsnAspLeuSerValLysA 277
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
632 CCAATAACCGCGAAAGTTTCATATCATATTCGAAGCGCATAT...TCTTGG 678
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
277 laAspAsnIleGlnThrAspSerGlnVal.....ArgTyrAspSerTyr 291
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
679 CTGCTCGGTGGCAATACCTTTGCACAAATGATCAGGTGGTGGCACAGT 728
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
292 AspLysAspGlySerGluAsnTyrGlnAsnTyrArgGlyGlyIleThrVa 308
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
729 C.....AACTTAGTACGCGAANA..... 747
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
308 laAsnAsnSerGlySerSerGlnThrLeuThrLysThrGluLeuLysGlyL 325
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
748 .....ATTAAACATAGCCCATATGTTTTCACCAACAGCAGGCG 786
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
325 ysAsnIleThrLeuValAlaSerSerHisAsnGlnIleLysAlaSerAsp 341
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
787 TCATTTCGGCAC.....AGTGGCTCACCATGTTTATC..... 819
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
342 LeuMetGlyAspIleThrLeuGlnGlyAlaAspLeuThrIleAspGl 358
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
819 ..... 819
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
358 yLysGlnLeuGlnGlnLysGluThrAspIleAspAsnArgTrpPheTyrS 375
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
820 .....TATGATGCCCAAGCAAGAGTGGTTAATTAATGGGTATTG 861
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
375 erTrpLysTyrAspValThrLysGluLys.....GluGlnIleGln 388
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
862 CAAACAGGCAACCCCTATATAGAAAGCAATGGCTTCCAGCTAGTTCG 911
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
389 GlnIleGlySerGlnIleAspAlaLysAsnAsnAlaThrLeuThrAlaTh 405
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
912 TAA.....GATTGGTCTATGAT.....GAAATCTTTGCTGGAGATACC 952
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
405 rLysGlyAspValThrLeuAspAlaAlaLysIleAsnAlaGlyAsnAl 422
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
953 ATTCAGTA.....TTCTACGACACCATCAAAATGGG.....AAA 987
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
422 euAlaIleAsnAlaAsnLysAspIleHisIleAsnGlyLeuValGluLys 438
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
988 TACTTTTTTAACGACAATAAAT..... 1011
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
439 GluSerArgSerGluAsnGlyAsnLysArgAsnHisThrSerArgLeuGl 455
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1011 ..... 1011
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
455 uSerGlySerTrpSerAsnSerHisGlnThrGluThrLeuLysAlaSerG 472
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1012 .....GGCGCAGGAAAA.....ATCGATGCCAAA..... 1035
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
472 luteuThrAlaGlyLysAspLeuGlyLeuAspAlaGlnGlySerIleThr 488
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1036 .....CATAAACACTATTCTTACCTTATAGATATAA 1067
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```



```
489 AlaGlnGlyAlaLysLeuHisAlaAsnGluAsnValLeuValAsnAlaLy 505
1068 AACAGCA.....ACGTTCAATGTTTAATGTTCTTTATCCGAGA 1108
505 sAspAsnIleAsnLeuAsnValGlnLysThrAsnAsnAspLysThrValT 522
1109 CAGCAAGAGAACCTGTTTATCATCTGTCAGGTGGGTCAACAGTTATCGA 1158
522 hrAspAsnHisValMetTrpGlyGlyLeuGly.....GlyGlnAsn 536
1159 CCCAGTGAATATGAGAGAAATATTTCTTTTATTGACAAA.....GG 1202
537 LysAsnAsnAsnAsnGlnGlnValSerHisAlaThrGlnLeuThrAl 553
1203 AARAGTGAATGATATCTACAGCACATCAACCAAGGCGCGGGT 1252
553 aAspGlyGlnLeuLeuLeuAlaAlaAspAsnAsnValAsnIleThrGlys 570
1253 TGTATTTTGAGGTAAATTTTACGGTCTCGCTAAACAAACGAAACGTGG 1302
570 erGlnValLysGlyAsn..... 575
1303 CAAGCGCGCGGCTTCATATCATGTCAGTACGTCACGTTACTTGGAAAGT 1352
576 GlnGlyAlaPheValLysThrThrGlnGlyAspValVal.....11 589
1353 AAACGGCGTGCACCAACACCGCTGTCACAAATC.....GGCA 1390
589 eAspAsnAlaLeuSerGluThrLysSerHisLysAsnGluThrAsnLysGln 622
1441 GTGGCGCAGCGTAAAGTCATCTTAGATCAGCAGCGGACGATCAAGGCAA 1490
623 ThrSerThrGlySerGluLeuIleSerAsp..... 632
1491 AARACAGCCTTTAGTCAATCGCTGTGTCAGCGCAGG..... 1530
633 .....AlaGlnLeuThrValValSerHisLysAsnValAsnVal 645
1531 .....GGCAGGTGCACACTGAATCCCGAT..... 1554
645 alIleGlySerLeuIleLysSerAlaAspLysLeuGlyIleHisSerLeu 661
1555 .....AATCAGTTCACCCCGA 1571
662 GlyAspIleAsnValLysSerAlaGlnGlnValThrLysIleAspAspGl 678
1572 CAACACTTATTCGGCTTTCCGGCGGACGTTTGGATTTCACAGGG..... 1617
678 uLysThrSerLeuAlaIleThrGlyHisAlaLysGluValGluAspLysG 695
1618 ..CATTCGCTTTCGTCACCGCATTCAAAATACCGATGAAGGGCGGATG 1665
695 lnTySerAlaGlyPheHis..... 701
1666 ATTCTCAACCAATCAAGACAAAGATCAACGCTTACCATTTACAGGCAA 1715
702 ...IleThrHisThrThrAsnLys...AsnThrSerThrGluThrGluGl 716
1716 TAAGATATTACTACACCGCATACACACTGGTAGCAAAAAG 1765
716 nAlaAsnSerThrIleSerGlyAlaAsnValAspLeuGlnAlaAsnLysA 733
1766 AATGTGCTCAACGGTGTGTTGGCGAGAAAGATCAACCAACAAACGAA 1815
733 spValThrPheAlaGly.....SerAspLeuLysThrThrAla 745
1816 GGGCGGCTCAATCTG..... 1830
```

---

```
746 GlyAsnAlaSerIleThrGlyAspAsnValAlaPheValSerThrGluAs 762
1831 .AATTACCAACCGGAAGACGGATCGCACTTTACTGCTTTCCGCGGAA 1879
762 nLysLysGlnThrAspAsnThrThr.....IleSerGlyGlyp 777
1880 CAAATTTAAACGGCAATATATCAGCAACAAACGGCAAACTGTTTTCAGC 1929
777 heSerTyThrGlyGlyValAspLysValGlySerLysAlaAspPheGln 793
1930 GGCAGACCCGACACCGCACCTTACATCATTTAGGAAGCGGTGTCATAA 1979
794 .....TyraPlysglnHisThrGlnThrGluValThrLy 805
1980 AATGGAAGTATCCCAACAGGAGAAATCGTGTGGAC..... 2016
805 sAsnArgGly...SerGlnThrGluValAlaGlyAspLeuThrIleThrA 821
2017 ..AACGATTGGATCGACCGCACATTTAAAGCGGAAACTTCCATATTCAG 2064
821 laAsnLys.....AspLeuLeuHisGluGlyAlaSerHisHisValGlu 835
2065 GCGGACAAAGCGGTGTTTCCGCAATGTTGCCAAAGTGAAGCGGATTG 2114
836 GlyArgTyThrGlnGluSerGlyGluAsnIleGln..... 846
2115 GCATTTAAGCATCAGCCCGCAAGCATTTTCGGTGTCCGACCGCATCAA 2164
847 .HisLeu.....AlaValAsnAsps 853
2165 GCACACCAATCTGTACAGTTCG..... 2187
853 erGluThrSerLysThrAspSerLeuAsnValGlyIleAspValGlyVal 869
2188 .....GACTGGACGGGTCTGACAAAGTGTACCGAAACCAATTCACGA 2231
870 AsnLeuAspTyThrSerGlyValThrLysProValLysLysAlaIleGluAs 886
2232 CGAT...AAGTGTATTGCTTCATTGACGAAGACCCAGCATCAGAGCAATG 2278
886 pGlyValAsnThrThrLysProGlyAsnAsnThrAspLeuThrLysLysV 903
2279 TCAGCGCTTCCCGATCAC...GCTCATTTA...AATCTCACAGGACTTGC 2322
903 alThrAlaArgAspAlaIleAlaAsnLeuAlaAsnLeuSerAsnLeuGlu 919
2323 ACACTCACAGCGCAATCTTAGTCAGGC..... 2349
920 ThrProAsnValGlyValGluValGlyIleLysGlyGlySerGlnGl 936
2349 ..... 2349
936 nSerGlnThrAspSerGlnAlaValSerThrSerIleAsnAlaGlyLysI 953
2350 .....GGAGACACGACCATATACG 2367
953 leAspIleAspSerAsnAsnLysLeuHisAspGlnGlyThrHisTy 968
2368 GTTACGCGCAACGCCCAACAAACGCAACCTCAGCTCGTGGGCAATGC 2417
969 .....GlnSerThrGlnGluGly...IleSerLeuThrAlaAsnTh 981
2418 C.....CAAGCAACATTTAATCAAG 2437
981 rHisThrSerGluAlaThrLeuAspLysHisGlnThrPheHisGluT 998
2438 CCACATTTAAACGGCAACACATCGCTTCGGACAACTGCTTCAATTAATCTA 2487
998 hrLysGlyGlyGlnIleGlyValSerThrLysThrGlySerAspIle 1014
2488 AGCACAACGCGCTACAAACGCGCAGTCTGAGCTTTCGACACAGCT... 2535
1015 ThrValAlaIleLysGlyGluGlyGlnThr.....AspAsnAlaLe 1029
```

2536 .....AAGCAACGTAAGCCATTCCCACTCAACGCAATGCT 2575  
1029 uMetGluThrLysAlaLysGlySerGlnPheThrSerAsnGlyAspIleS 1046  
2576 CCTAGCCGATAAGGCACTATTCCATTGTAACAGCGCTTT 2619  
1046 erIleAsnValGlyGluAsnAlaHisTyrGluGlyAlaGlnPheAspAla 1062  
2620 ....ACCGGAAAA.....ATCAGCGCGGCAAGGATACGGCATTACACTT 2660  
1063 GlnLysGlyLysThrValIleAsnAlaGly..... 1072  
2661 AAAACACAGCAGATGAGCGCTGCCCTCGGCACGGAATTAGCAATTAA 2710  
1073 .....GlyAspLeuT 1076  
2711 ACCTTGACAAGCCACCATTACACTCAATTCCGCCATTCGACACAGATCG 2760  
1076 hrLeuAlaGlnAlaThr..... 1081  
2761 GCAGCGCGCAAAACCGGAGTTCGGCAGATGCGCGCGCGCTTCGCG 2810  
1081 ..... 1081  
2811 CGTTCCCTATTATCCGTTAGCCGCCCAACTTCGGCAGAAATCCCGTTTCA 2860  
1082 .....AspThrHisSerGluSerGlnSerA 1090  
2861 ACACGCTGACGTTAAACGGAATTTGAAGTCGACGAGGCAACATCCGCTTT 2910  
1090 sn.....ValasnGlySerAlaAsn..... 1096  
2911 ATGTCGGAACCTTCGGCTACCGCAGCGGCAAAATTTGAGCTGGCGGAAAG 2960  
1097 .....LeuLysValGlyThrTh 1102  
2961 TTCGAAAGGCACTTACACTTGGCTGTCTCAACAATACCGGCAACGACCG 3010  
1102 rProGluSerLysAspTyrGlyGlyPheAsnAlaGlyThrThrHisH 1119  
3011 TAAGTCTCGAGCAATTTAGC.....GTAGTGGAGGAAAGAC 3048  
1119 isSerLysGluGlnThrThrAlaLysValGlyThrIleThrGlySerGln 1135  
3049 AACACACGCTGTCCGAAATCTTAATTCACCTGCAAAACGAGACGT 3098  
1136 GlyIleGluLeuAsnAlaGlyHisAsnLeuThrLeuGlnGlyThrHisLe 1152  
3099 CGATGCCGCGCATGGCGTTATACGTTATCCGCAAGACGGCGAGTTCC 3148  
1152 u..... 1152  
3149 GCCTGCATAATCCGCTCAAGAACAAGAGCTTTCGGACAAATCGGCAAG 3198  
1152 ..... 1152  
3199 CGGGAGAAACAGAGCGCGCTTGACGGCAAAA.....CAGGCACA 3239  
1153 SerSerGluGlnAspIleAlaLeuAsnAlaThrAsnLysValAspLeuG 1169  
3240 ACTTGCGCCCAACACACAGCGGGAAGAACACACGCGCCCAAGCCTTCAGC 3289  
1169 nSerAlaSerSerGluHisThrGluLysGlyAsnAsnLeuSerGlyGly 1186  
3290 CGCTGATTGCGCGCGGCAATGCCACCGGAAAGGAGCAAGAGTGT... 3336  
1186 alGlnAlaGlyPheGlyLysLysMethThrAspAlaSerSerValAsn 1202  
3337 .....GCCGAACCGGC 3347  
1203 GlyLeuGlySerAlaGlnPheAlaIleGlyLysGlnAspGluLysSerVa 1219

seq\_name: pir2.S15204

3348 CCGCAGGCAAGCGGG.....GAATATCGCGGCATTTATCCAGCGGAGG 3391  
1219 lSerArgGluGlyGlyThrIleAsnAsnSerGlyAsnLeuThrIleAsnG 1236  
3392 AAGAGAAAAACGGGTGACGGG.....GATAAAGACACCGCC 3429  
1236 lYsnSerValHisLeuGlnGlyAlaGlnValAsnSerLysAspThrGln 1252  
3430 TTGGCGAAGACAG..... 3441  
1253 LeuThrSerGlnSerGlyAspIleGluIleThrSerAlaGlnSerThrAs 1269  
3441 ..... 3441  
1269 pTyrLysAsnAsnTrpGlyThrAspIleGlyPheAsnGlyLysLysThra 1286  
3442 .....CGGAAGCGGAAACCGCGCGGTACCACCGCTTC 3477  
1286 snAsnThrProLysGluValThrGluGluLysProAlaThrSerIleHis 1302  
3478 CCGCGCGCGCGCGCGCGGATTTCGCCGAACCGCAGCGCCCAACC 3527  
1303 AsnIleGlyLysLeuValAsnValGluAspGlnGlnLys..... 1317  
3528 GCAACCCCAACCGCAGCGGACCTGATCAGCCGTTATGCCAATAGCGTT 3577  
1318 .....ThrSerHisGlnAsnAlaThrL 1325  
3578 TGAGTGAATTTCCCGCAGCTCAACAGCGTTTTCGCCGTACAGGACGAA 3627  
1325 euGluThrGlyThrLeuThrIleAsnSer.....AsnLysAsp 1337  
3628 TTGGACCGCGCTGTTCCGGAAGACCGCGCAACCGCGTTTGGACACAGCG 3677  
1338 LeuThrLeuSerGlyAlaAsnValThrAlaAspSerVal...ThrGlyAs 1353  
3678 CATCCGGGACACCAACACTACCGTTTCGCAAGATTTCCGCGCTACCGCC 3727  
1353 nValGlyGlySerLeuAsnIleAlaSerGln.....L 1364  
3728 ACAAACCGACCTGCGCCAAATCGGTATCCAGAAAACCTCGGCAGCGG 3777  
1364 yGluSerAsp...ArgHisVal..... 1370  
3778 CGGTCGCGCATCCTGTTTCGCACACCGCGCGGAAACACCTTCGACGA 3827  
1371 ThrValGlyValAsnValGlyTyrAsnHisThrAsnAspProLysSerSe 1387  
3828 CGGCATCGGCACTCGCGACCGGTTGCCCGCGCTTCGGTTCGGGCAAT 3877  
1387 rGlnValAsnLysThrAlaLys...AlaGlyGlySerLeuLeuGluLysT 1403  
3878 ACGGCATCGCAGGTTTCGACATCGCATCGCGCGCGCGCGGTTTACT 3927  
1403 hrIleLysAspThrIleAspSerGlyIleLysSerSerThrAspAlaIle 1419  
3928 AGC.....GGCAGCCTTTTCAGACGCGCATCAGAGCAAAATCCGCGC 3968  
1420 SerAspLysTyrAsnSerLeuSerSerThrIleAlaAspLysThr..... 1434  
3969 CCGCGTGTGCTTACGGCATT.....CAGGCAAGATACCGG 4006  
1435 .....GlyIleSerAspGluThrLysAlaLysIleAspG 1446  
4007 CAGGTTTCGCG.....GGATTCGCGCATCGAACCGCACATCGCGCAACG 4050  
1446 lngLysPheGlyLysValGlyAsnGlyIleLysAsnIleValThrGlyAla 1462  
4051 CGCTATTTCGTCCTCAAAAGCGGATTACCGATACGAAACGTCAT 4095  
1463 GluGlyHisThrAlaAsnAlaAspIleLysValThrHisValAsp 1477

## seq documentation\_block:

pertactin - Bordetella parapertussis  
N:Alternate names: outer membrane protein P70  
C:Species: Bordetella parapertussis  
C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 08-Oct-1999  
C:Accession: S15204; S14659  
R:Li, L.J.; Dougan, G.; Novotny, P.; Charles, I.G.  
Mol. Microbiol. 5, 409-417, 1991  
A:Title: P.70 pertactin, an outer-membrane protein from Bordetella parapertussis: cloning  
A:Reference number: S15204; MUID:91251771  
A:Accession: S15204  
A:Molecule type: DNA  
A:Residues: 1-922 <LIL>  
A:Cross-references: EMBL:X54547; NID:g39761; PIDN:CAA38419.1; PID:g39762  
C:Genetics:  
A:Gene: prn  
C:Keywords: membrane protein

## alignment\_scores:

Quality: 275.50 Length: 1162  
Ratio: 0.566 Gaps: 54  
Percent Similarity: 41.910 Percent Identity: 21.687

## alignment\_block:

US-09-303-518D-653 x S15204 ..

Align seg 1/1 to: S15204 from: 1 to: 922

1366 AACGACGGCGCTGCCAAAATCGCAAGGC.....ACGCT 1400  
|||||  
2 AsnMetSerLeuSerArgIleValLysAlaProLeuArgThr 18  
1401 GCTGGTTCACCCAAAGG.....  
|||||  
18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaProAlaAlaTyrAlaA 35  
1420 .....GAAACCAAGGCTCGTCAAGCTGGGCGACGTAAG...GTCATC 1461  
|||||  
35 spTrpAsnAsnGlnSerIleLeuLysAlaGlyGluArgGlnHisGlyIle 51  
1462 TTGATCAGCAGCGCGCGATCAAGCAAAAACAGCCTTTAGTGAAT 1511  
|||||  
52 HisIleLysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThr 68  
1512 CGGTGTCAGCGCGAGGGGAGCGTCAACTGAATGCCGATAATCAGT 1561  
|||||  
68 rIleLysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn... 83  
1562 TCACCCCGCAAACTCTATTTCGGCTTCGCGCGGACGTTGGATTG 1611  
|||||  
84 .....ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSer 97  
1612 AACGGCGCATCGCTTCCTCCCGCATTCAAAATACCGATGAAGGGGC 1661  
|||||  
98 SerGly.....GlnLeuPheAspGluGlyVal 106  
1662 GATGATTGTCAACCAACATCAAGCAAGAATCCACCGTTACCATACAG 1711  
|||||  
106 l.....ArgAlaPheLeuGlyThrValThrValLysA 117  
1712 GCATAA.....GATATTACTACACCGCAATACACAACTTG 1752  
|||||  
117 laGlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValSerAspThr 133  
1753 GATAGCAAAAGAAATGGC...TACAACGGTTGGTTTGGCGAGAAGA 1799  
|||||  
134 ArgAspAspGlyIleAlaLeuTyrValAla.....GlyGluGln.. 147  
1800 TGCAACCAAAACGAAGCGGCGCTCAATCTGAATTACCAACCGGAAGAAG 1849  
|||||  
148 .....AlaGlnAlaSerIleA 153

1850 CGSATCGCACCTTTACTGCTTTCGCGCGAACAATAATTA.....AACGGC 1893  
|||||  
153 laAspSerThrLeuGlnGlyAlaGlyValArgValGluArgGlyAla 169  
1894 AATATCACCAACAAAGCGCAAACTGTTTTCAGCGCGACAGCACACC 1943  
|||||  
170 AsnValThrValGlnArgSerThrIleValAspGlyGlyLeu..... 183  
1944 GCACGGCTACATCATTTAGGAAGGGGTGTC...AAATGGAAGTA 1990  
|||||  
184 .....HisIleGlyThrLeuGlnProLeuGlnProGluAspL 196  
1991 TCCCAAGGAGAAATCGTGTGGGCAACAGATTGGATCGACCGC..... 2034  
|||||  
196 euProProSerArgValValLeuGlyAspThrSerValThrAlaValPro 212  
2035 .....ACATTTAAAGCGGAAACTTCCA 2057  
213 AlaSerGlyAlaProAlaAlaValPheValPheGlyAlaAsnGluLeuTh 229  
2058 TATTACGGCGGACAGCGGTGTTCCCGCAATGTTGCCAAAGTGGAG 2107  
|||||  
229 rValAspGlyGly..... 233  
2108 GCGATTGGCATTTAAGCAATCACGCCAAGCAGTTTCGGTGTGCGACCG 2157  
|||||  
234 .....HisIleThrGlyGlyArgAlaAla.....GlyValAla... 244  
2158 CATCAAGCCACACAATCTGTACAGTTCGGACTGGACGGGTCTGACAG 2207  
244 ..... 244  
2208 TTGTACCGAAAAACCATTTACGACGATAAAGTATTGCTTCTATTGACCA 2257  
|||||  
245 .....AlaMetAspGlyAlaIleValHisLeuGlnA 255  
2258 AGACGACATCAGAGGCAATGTGAGCTTGGCGCATCAGCTCATTTAAAT 2307  
|||||  
255 rGalaThrIleArg..... 259  
2308 CTCACAGGACTTGCACACTCAACGCCAATCTTAGTGCAGCGGAGACAC 2357  
|||||  
260 .....ArgGlyAspAlaProAlaGlyGly..... 267  
2358 GCACTATACGTTTACCGCAACGCCACCAACAAAGCGCACCTC..... 2400  
|||||  
268 .....AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282  
2401 .....AGC 2403  
282 lyPheGlyProLeuLeuAspGlyTyrTyrGlyValAspValSerAspSer 298  
2404 CTGCTGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAACGGCAA 2453  
|||||  
299 ThrValAspLeuAlaGlnSerIleValGluAlaProGlnLeuGlyAlaAl 315  
2454 CACATCGCTTCGGACAATGCTTCATTTAATCTAAGC..... 2490  
|||||  
315 aIleArgAlaGlyArgGlyAlaArgValThrValSerGlyGlySerLeuS 332  
2491 .....AACACGGCGTACAAAACGGC..... 2511  
|||||  
332 erAlaProHisGlyAsnValIleGluThrGlyGlyAlaArgArgPhe 348  
2512 .....AGTCTGACGCTTTCGACACAGCTAAGGC 2540  
|||||  
349 ProProAlaSerProLeuSerIleThrLeuGlnAlaGlyAlaArgAl 365  
2541 AAAC.....GTAAGCCATT 2554  
|||||  
365 aGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProValLysLeuT 382



846 alGlnProTyrIleLysAlaSerValLeuGlnGlupheAspGlyAlaGly 862  
4231 CGAAGCGCGTCATATACCGCGGTATTGGCGAGGATTTGGCAAAACCCG 4280  
:::|||||::: |||  
863 ThrValArgThrAsnGlyIleAlaHisArgThrGluLeuArgGlyThr 879  
4281 CAGTGGCGAATGGCGCTAACCGCCGAAATCAAGGTTTCACGCTGTCCC 4330  
I ||||| |||  
879 g...AlaGluLeuGly.....LeuGlyM 886  
4331 TCCACGCTGCGCGCCCAAGCGCGCAATTTGGAAGCG.....CAGCAC 4374  
::: |||||::: ||| |||  
886 etAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGluTyr 902  
4375 AGCGCGGCATCAATTTAGCTACCGCTGG 4404  
||| ||| |||||::: |||  
903 SerLysGlyProLysLeuAlaMetProTyr 912

seq\_name: pir2:M47675

seq\_documentation\_block:

68K outer membrane protein P.68 pertactin - Bordetella bronchiseptica

C:Species: Bordetella bronchiseptica

C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999

C:Accession: A47675

R:Li, J.; Fairweather, N.F.; Novotny, P.; Dougan, G.; Charles, I.G.

J. Gen. Microbiol. 138, 1697-1705, 1992

A:Title: Cloning, nucleotide sequence and heterologous expression of the protective oute

A:Reference number: A47675; MUID: 92407514

A:Contents: CN7531

A:Accession: A47675

A>Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-911 <L1>

A:Cross-references: GB:X54815; GB:S46416; NID:g39396; PIDN:CAA38584.1; PID:g39397

A>Note: sequence extracted from NCBI backbone (NCBIN:113318, NCBI:P:113319)

alignment\_scores:

Quality: 274.00 Length: 1149

Ratio: 0.584 Gaps: 50

Percent Similarity: 40.818 Percent Identity: 20.540

alignment\_block:

US-09-303-518D-653 x A47675 ..

Align seg 1/1 to: A47675 from: 1 to: 911

1306 GCGCGCGCGTTCATATACGATGCGAGTACCGTACTTGGAAAGTAA 1355  
|||||::: |||::: |||  
58 GlyAlaGlyValArgThrAlaThrGlyThrIle.....LysValSe 72  
1356 CGGCGTGGCAACGACCGCTGTCCAAATCGCAAGGACCGCTGCTGG 1405  
||| ::|::|::|::|::|  
72 rGly.....ArgGlnAlaGlnGlyValLeuLeug 82  
1406 TTCAAGCCAAAGGGAAACCAAGGCTCGCTGCGGTGGCGACGTTAA 1455  
::: |||::: |||  
82 LuAsnProAlaAlaGlu.....LeuArgPheGlnAsnGlySer 94  
1456 GTCATCTTAGATCAGCAGCGCGAGATCAAGGCAAAACACGCTTTAG 1505  
||| ::|::|::|::|::|  
95 ValThrSerSerGlyGlnLeuPheAspGluGlyValArgPheLeu.. 110  
1506 TGAATCGGCTGTGTCAGCGCGAGGGGCGGTGCAACTGAATGCCGATA 1555  
|||::|::|::|::|  
111 .....GlyThrValThyValLysAla.... 117  
1556 ATCAGTTCAACCCGACAAACTCTATTTCGCTTCGCGCGCGAGCTTG 1605  
|||::|::|::|::|  
118 .....GlyLysLeu 120  
1606 GATTGTAACGGGATTCGCTTTCGTTCCACCGCATTCAAAATACCGATGA 1655

308 GlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrValSerGlyCl 324  
2491 .....AACACCGCGGTACAAAACGCG..... 2511  
324 ySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyAlaAla 341  
2512 .....AGTCTACGCTTCCGACAAAC 2532  
341 rgArgPheProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357  
2533 GCTAAGGCAAC.....GT 2546  
358 AlaArgAlaGlnGlyArgAlaLeuLeuTyArgValLeuProGluProVa 374  
2547 AAGCCATCCGACCTACAGGCAATGTCTCCCTAGCCGATAGGCAAT 2596  
374 llyLeuThrLeuAlaGlyGlyAlaGlnGlnGlyAsp.....IleV 389  
2597 TCCATTTGAAACACGCGCTTACCGGAAATACAGCGCGGCAAGAT 2646  
389 alAlaThrGluLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405  
2647 ACGGCATTACACTTAAAGACAGCGAATGAGCGTCCGCTGGGCGACGGA 2696  
406 ValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaThrArgAl 420  
2697 ATTAGCAATTTAACTTGACAGCGCAACACCACTTACACTCAATCCGCT 2746  
420 aValAspSerLeuSerIleAspAsnAlaThrTrpValMetThrAsp.... 435  
2747 ATCGACACGATCGCGAGCGCGCAACCGGCGAGTGGCGGACGCGCG 2796  
436 .....AsnSerAsnValGlyAlaLeu 442  
2797 CCGCGCGGTTCGCGCGTTCCTATTATCCGTACGCGCGCAACTTCGCG 2846  
443 ArgLeuAlaSerAspGlySer...ValAspPheGlnGlnProAlaGluAl 458  
2847 AGAATCCGCTTACACGCTGACGCGTAACGGCAATTTGACGCTCAGG 2896  
458 a...GlyArgPheLysCysLeuMetValAsp...ThrLeuAlaGlySerG 473  
2897 GAACATTCGCTTTATGTGCGAACTCTTCGGCTACCGCAGCGGCAAAATG 2946  
473 llyLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeu 489  
2947 AAGCTGGCGGAAAGTTCGAAAGGCACTTACACCTTGGCTGTCAACAATAC 2996  
490 ValValMetArgAspAlaSerGlyGlnHisArgLeuLeuValArgAsnSe 506  
2997 CGGCAACGAAACCGTAAGTCTCGGACCAATTCAGCGTAGTGAAGAAAG 3046  
506 rGlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln..... 520  
3047 ACAACACACCGCTGTCCGAAATCTTAATTTACACCTGCAAAACGAA... 3093  
521 .....ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAsp 535  
3094 ...CACGTCGATCGCGCGCTTATCAGCTTATCCGCAAA...GA 3137  
536 GlyLysValAspIleGlyThrTyArgTrpArgLeuAlaAlaAsnGlyAs 552  
3138 CGCGAGTTCGCTGCATATTCGGTCAAGAACACAGAGCTTTCGAC 3187  
552 nGlyGlnTrpSerLeu.Val..... 558  
3188 AACTCGGCAAGCGGAGAAACAGAGCGCGCTTTCACGGGCAAAACAGGCA 3237  
558 ..... 558  
3238 CAATTCGCGCAACAAACAGCGGAAAGACAAACGCGCAAGCCTTGA 3287  
558 ..... 558

3288 CGGCTGATTTCGGCGCGCGCAATCCACGAAAGGCAAGTGTTC 3337  
559 .....GlyAlaLysAlaProProAlaProLysProAla 569  
3338 CCNAACCGCGCGCAGCGAGGGGAAAATGCCGCAATTATTCAGGCG 3387  
570 Pro...GlnProGlyProGln..... 575  
3388 GAGGAGAGAAACCGGTGCAGGCGGATAAAGACACCGCTTGGCGAA 3437  
575 ..... 575  
3438 ACAGCGGAAAGCAACCGCGCGGTACACCGCTTCCCGCGCGCC 3487  
576 .....ProGlyProGlnProProGln.ProProGlnPr 586  
3488 GCGCGCGCGCGGTATTCGCCAACCAGCGAGCCCAACCGCAACCCCA 3537  
586 oProGlnProProGlnArgGlnProGluAla.ProAlaProGlnProPro 602  
3538 CCGCAGCGCGACCTGATCAGCGCTTATGCCAATAGCGTTTGTAGTGAAT 3587  
603 AlaglyArgGluLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrGl 618  
3588 TTCGCGCGCGCTCAACAGCGTT.....TTCGCGGTACAGGACGAATTGG 3631  
618 yGlyValGlyLeuAlaSerThrLeuTrpTyArgAlaGluSerAsnAlaLeuS 635  
3632 ACCGCGTGTTCGCGAGACCGCGCGCAAC.....GCCGTTGG 3669  
635 erLysArgLeuGlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTrp 651  
3670 ACAAGCGCGCATCGGGGACACCAACACTACCGTTTCGCAAGATTTCGCGCG 3719  
652 GlyArgGlyPhe.....Al 656  
3720 CTACCGCGCAACACCGACCTGCGCCAAATCGGTATGCAGAAACCTCG 3769  
656 aGlnArgGlnGlnLeuAsp..... 662  
3770 GCAGCGCGCGTTCGCGCATCTCTTTTCGCAACACCGCGGAAACACC 3819  
663 .....AsnArgAlaGlyArgArg 668  
3820 TTCGACGAC.....GGCATCGGCAACTCGCGCAGC 3848  
669 PheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisAlaValAl 685  
3849 GCTTGGCGCGCGTGC.....GTTTTCGGGCAATACGGCATCGGCGAGT 3892  
685 aValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTyThrArgG 702  
3893 TCACATCGCATCAGCGCGCGGTTTAGTAGCGGCGAGCTT... 3939  
702 llyAspArgGlyPheThrGlyAspGlyGlyHisThrAspSerValHis 718  
3939 ..... 3939  
719 ValGlyGlyTyAlaThrTyIleAlaAsnSerGlyPheTyLeuAspAl 735  
3940 .....TCAG 3943  
735 aThrLeuArgAlaSerArgLeuGluAsnAspPheLysValAlaGlySerA 752  
3944 ACGGC.....ATCAGAGGCAAAATCCGCGCGCTGCTGCTTACGCG 3987  
752 spGlyTyAlaValLysGlyLysTyArgThr.....HisGly 764  
3988 ATTCAGGCAAGATACCGCGAGT.....TTCGGCGGAT 4022  
765 ValGlyAlaSerLeuGluAlaGlyArgPheAlaHisAlaAspGlyTr 781



326 lysGly..... 328  
940 GCTGGAGATACCATTCAGTATTCTACGAACACATCAAAATGGGAATA 989  
329 ...GlyAspGlyIleGlyPheAlaPheSerProGlyValIleuGlyCyluTh 344  
990 CTTTTTAACGCAATAATAATGGCCAGGA..... 1020  
344 rGlyLeuAsnGlyAlaAlaValIleGlyGlyLeuSerAsnAlaPheG 361  
1021 ....AAATCGATGCCAACAACATAAATCTATTCTACCTTATAGATTA 1065  
361 lyPheLysLeuAspThrTyrHis...AsnThrLysProAsn..... 374  
1066 AAAACACGAACCGTTCATATGTTTAAATGTTCTTTCACGACACGAAG 1115  
375 .....SerAlaAlaLysAlaAsnAl 381  
1116 AGAACCTGTTTATCGTGCAGGTGGGTCAACAGTATTCACCCAGAC 1165  
381 aspProSerAsnValAlaGlyGlyAla..... 391  
1166 TGAATAATGGAGAAATATTCCTTTATTGACAAAGGAAAGTCAATTG 1215  
392 ....PheGlyAlaPheValThrThrAspSerTyrGlyValAlaThrThr 406  
1216 ATACTTACGACCAATCAACCAAGCGCGGCGTGTGTTTATTTTGGAGG 1265  
407 TyrThrSerSerThrAlaAspAsnAlaLysLeu..... 419  
1266 TAATTTTACGGTCTCGCTTAAACCAACGAACGAGTGGCAAGCGCGCGC 1315  
420 .....AsnValGlnProThrAsnAsn...ThrPheGlnAspPheAspI 433  
1316 TTCATATCAGTATGCGAGTACCGTT..... 1341  
433 leAsnTyrAsnGlyAspThrLysValMetThrValLysTyrAlaGlyGln 449  
1342 ACTTGGAAAGTAAAGCGGTGGCAACGACCGCTGTCCAAATCGGCAA 1391  
450 ThrTrpThrArgAsn....IleSerAspTrpIleAlaLysSerGlyTh 464  
1392 AGGCACGCTGCTGTCTCAAGCAAGGGGAAACCAAGCTCGTCAAGC 1441  
464 rThrAsnPheSerLeuSerMetThrAlaSerThrGlyGlyAlaThrAsnL 481  
1442 TGGCGGACGGTAAAGTCATCTTAGATCAGCAGGCGGACGATCAAGGCAA 1491  
481 eu.....GlnGlnVal.....GlnPheGly 487  
1492 AAACAAAGCCTTTAGTAAATCGGTTGGTCAGCGGACGGGACGGTGCA 1541  
488 ThrPheGluTyrThrGluSerAlaValThrGlnValArg...TyrValAs 503  
1542 ACTGAATCCGATATCATCTCAACCCGCAAACTCTATTTCGGCTTC 1591  
503 pValThrThrGlyLysAspIleProProLysThrTyrSerGly.... 518  
1592 GCGCGGACGTTTGGATTTGAACGGGATTCGCTTTCACCGCAT 1641  
518 ..... 518  
1642 CAATAATCCGATGAAGGGCGGATGTTCAACCAATCAAGCAAGA 1691  
519 ...AsnValAspGlnValThrIleAspAsn.....GlnG 530  
1692 ATCCACGCTTACCATACAGGCAATAAGATATTACTACAACCGGCAATA 1741  
530 nSerAlaLeuThrAlaLysGlyTyr.....AsnT 540  
1742 ACAACAACCTCGATAGCAAAAGAAATGCTTCAACAGCGTTGGTTGGC 1791  
.....

540 yrThrSerValAspSerSerTyrAlaSerThrThrTyrAsnAsp.....Thr 554  
1792 CAGAAAGATGCAACCAACGACGGCGGTCAATCTCAATTAATAC..... 1836  
555 AsnLysThrValLysMetThrAsnAlaGlyGlnSerValThrTyrPhe 571  
1837 .....CAACCGGAAGAAG 1849  
571 eThrAspValLysAlaProThrValThrValGlyAsnGlnThrIleGluV 588  
1850 CGGATCGCCTTTA.....CTGCTTCCGCGGGAACAAATTTAAAC 1890  
588 alGlyLysThrMetAsnProValValLeuThrThrThrAspAsnGlyThr 604  
1891 GGCAATATCAGCAACAAACGCGCAAACTGTTTTCAGCGGACAGCCGAC 1940  
605 GlyThrValThrAsnThr.....ValThrGlyLeuProSe 616  
1941 ACCGACGCTTACATCAATTTAGGAACGGGTGTCAAAATGGAAGTA 1990  
616 rGlyLeuSerTyrAsp.....SerAlaThrAsnSerIleIleGlyT 630  
1991 TCCACAA.....GGAGAAATCGTGTGGGACACGATTCGATCGACCGC 2034  
630 hrProThrLysIleGlyGln.SerThrValThrValValSerThrAspG 646  
2035 ACATTTAAAGCGGAAACTTCCATATTTACAGGCGGAC.....AAGCGT 2078  
646 nAlaAsnLysSerThrThrThrPheThrIleAsnValValAspThrT 663  
2079 GGTTCGCCCAATGTTGCCAAAGTGAAGCGGATTCGCAATTTAAGCAAT 2128  
663 hrAlaProThrValThrPro.IleGlyAspGlnSerSerGluValTyrSe 679  
2129 AGCCCAAGCAGTTTCGGTGTGCGACCGCATCAAGCCACACAATCTGT 2178  
679 rProIleSerProIleLysIleAlaThrGlnAspAsnSerGlyAsnAlav 696  
2179 ACAGTTCGGACTGGAGCGGTCTGACAAGT.....TGATC 2213  
696 alThr.AsnThrValThrGlyLeuProSerGlyLeuThrPheAspSerTh 712  
2214 CGAAACCAATTAACGACGATAAAGTG.....ATTGCTT 2248  
712 rAsnAsnThrIleSerGlyThrProThrAsnIleGlyThrSerThrIles 729  
2249 CATTGAGCAAGCAGACATCAGAGGCAATGTCAGCCTTCCCGATCAGCT 2298  
729 erIleValSerThrAspAlaSerGlyAsnLysThrThrThrPhe... 744  
2299 CATTTAAATCTACAGGACTTGCCACACTCAAGGCAATCTTAGTCAGG 2348  
744 ..... 744  
2349 CGGAGACGCACTATACGGTTACGCGCAACGCCAACCCAAACCGCAACC 2398  
745 .....LysTyrGluValThrArgAsn...SerMetSerAspSerV 757  
2399 TCAGCCTCGTGGCAATGCCAA.....GCAACATTTAAT 2433  
757 alSerThrSerGlySerThrGlnGlnSerGlnSerValSerThrSerLys 773  
2434 CAAGCCACATTAACGCAACACATCGGCTTCGCAATGCTTCATTTAA 2483  
774 AlaAspSerGlnSerAlaSerThrThrSerGlySerIleValValSe 790  
2484 TCTAAGCAACAGCGGTACAAACAGGCGAGTCTGAGCTTCCGACACAG 2533  
790 rThrSerAlaSerThrSerLysSerThrValSerValSerLeuSerAspSerV 807  
2534 CTAAGCAACCACTAAGCCATTCGCACTCAAGGCAATGCTCCCTAGCC 2583  
807 alSerAlaSerLysSerLeuSerThrSerGluSerAsnSerValSerSer 823



3460 CCGGCTACACCGCCTTCCCGCGCCCGCGCCCGCGCGGATTTGCC 3509  
1067 GlnSerGlySerThr.SerThrSerThrSerAlaSerValArgT 1083  
3510 GCAACCGCAGCCCAACCGCAACCCCAACCGCAGCGACCTGATCAGCC 3559  
1083 hrSerGluSerGlnSerThrSerGlySerMetSerAlaSerGlnSerAsp 1099  
3560 GTTATGCCAATACGGTTTTCAGTGAATTTCCGCCACGCTCAACACGGTT 3609  
1100 SerMetSerIle.....SerThrSerPh 1107  
3610 TTCGCCGTACAGCAGAAATTTGGACCGCGTGTTCGCCAAGACCGCGCAA 3659  
1107 eSerAspSerThrSerAsp.....SerLysSerAlaSerT 1119  
3660 CGCGGTTTGGACAA.....CGGCATCCGGGACACCAAC 3694  
1119 hrAlaSerSerGluSerIleSerGlnSerAlaSerThrSerThrSerGly 1135  
3695 ACTACCGTTCGCAAGATTTCCGGCGCTACCGCCCAACAAACCGACCTCGC 3744  
1136 SerValSerThrSerThrSerLeuSerThrSerAsnSerGluArgThrSe 1152  
3745 CAAATCGGTATGCAGAAAAACCTCGCGAGCGCGCGCTCGGCATCTCTGTT 3794  
1152 rThrSerValSerAspSerThrSerLeuSerThrSerGluSerAspSerI 1169  
3795 TT.....CGCACACCGGACCGGAAACACCTTCGACGACGGCATCGCA 3838  
1169 leSerGluSerThrSerThrSerAspSerIleSerGluAlaIleSerAla 1185  
3839 ACTCGGCACGCTTCCCGCGCTGCTTTTCGGCAATACGGCATCGGC 3888  
1186 SerGluSerThrSerIleSerLeuSerGluSerAsnSerThrSerAspSe 1202  
3889 AGGTTCGACATCGCATCGCGCGCGGGTTCCTAGTAGCGCAGCGCT 3938  
1202 rGluSerGlnSerAlaSerAla.....PheLeuSerGluSerLeuS 1216  
3939 TTCAGACGGCATCAGAGCAAAATCCCGCGCGCGTCTGCTGATTCAGCA 3988  
1216 erGluSerThrSerGluSerThrSerGluSerValSerSerThrSer 1232  
3989 TTCAGGCAAGATACCGCGCAGTTTCGGCGGATTTCGGCATCGAACCGCAC 4038  
1233 GluSerThrSerLeuSerAspSerThrSerGluSerGlySerThrSerTh 1249  
4039 ATCGCGCAACCGCTATTTCGTCCAAAAGCGGATTACCGATACGAAAA 4088  
1249 rSerLeuSerAsnSerThrSerGlySerAlaSerIleSer...ThrSerT 1265  
4089 CGTCAATATCGCCACCGCGGCTTCATTCACCGCTACCGCGCGGC. 4137  
1265 hrSerIleSer.....GluSerThrSerThrPheLysSer 1276  
4138 .....ATTAGCGCAGATTATTTCATCAACCGCGCAACACATTTCCAT 4181  
1277 GluSerValSerThrSerLeuSerMetSerThrSerThrSerLeuSerAs 1293  
4182 CACGCGTTATTTCAGCGCTGCTCTATACCGATCCCGCTTCGCGCAAGTCC 4231  
1293 nSerThrSerLeuSerThrSerLeuSerAspSerThrSerAspSerLysS 1310  
4232 GAACGCGCTCAATACCGCGCTATTGGCGCAGGAT...TTCGGCAAAACC 4278  
1310 erAspSerLeuSerThrSerMetSerThrSerAspSerIleSerThrSer 1326  
4279 CGCAGTGGCGAATGGCGCTAAACCGCGCAATCAAGGTTTCACG 4323  
1327 LysSerAspSerIleSerThrSerThrSerLeuSerGlySerThr 1341

seq\_name: pir2:c82199

2584 GATAAGCAGATATTCATTTTGAACAACAGCGCTTTACCGGAAAAATCAG 2633  
824 SerThrSerThrSerLeuValAsnSerGlnSerValSerSerMetSe 840  
2634 CGCGCGCAAGATACCGCATTTACACTTAAAGACAGCGCAATGGACGCTGC 2683  
840 rGlySerValSerLysSerThrSerLeuSerAspSerIleSerAsnSerA 857  
2684 COTCGCGCACGGAATTAGCAATTTAAACCTTTGACACGCCACCATTTACA 2733  
857 snSerThrGluLysSerGluSerLeuSerThrSerThrSerAspSerLeu 873  
2734 CTCATTCGCGCTATCGACAGATCGCGCAGCGCGGCAAAACC...GGCAG 2780  
874 ArgThrSerThrSerLeuSerAspSerLeuSerMetSerThrSerGlySe 890  
2781 TCGCGCAGATCGCGCGCGCGCTTCGCGCGCTTCCTATATTCGCGTA 2830  
890 rLeuSerLysSerGlnSerLeuSerThrSerIleSerGlySerSerT 907  
2831 CGCGCGCAACTTCGGCAAAATCCCGTTTCAACACGCTGACGTAACCGC 2880  
907 hrSerAlaSerLeuSerAspSerThrSerAsnAlaIleSerThrSerThr 923  
2881 AAATTGAACGGTCAGGGAACATTCGCGTTTATGTCGGAACCTTCGCGCTA 2930  
924 SerLeuSerGluSerAlaSer..... 930  
2931 CGCGACGCAAAATCAAGCTGGCGAAAGTTCCGAAGCACTTACACCT 2980  
931 ThrSerAspSerIleSerIleSerAsnSerIleAlaAsnSerGlnSerA 947  
2981 TGCGTCTCAACATACCGCACAGACCGCTAAGTCTCGAGCAATTCAGC 3030  
947 laSerThrSerLysSerAspSerGlnSerThrSerIleSerLeuSerThr 963  
3031 GTAGTGAAGGAAGAACACACACCGCTGCTCGGAAATCTTAATTCAC 3080  
964 .....SerAspSerLysSerMetSerThrSer..... 972  
3081 CTGCAACAGCAACAGTCGATCCCGCGCGCTATACGCTTATACCT 3130  
972 ..... 972  
3131 GCAAGACGCGAGTTCGCGCTGCATAATCCGCTCAAGAACAGAGCTT 3180  
973 .....GluSerLeu 975  
3181 TCGCACAACTCGGCAAGCGGAGAGAACAGAGCGCGCTTCAGCGCAAA 3230  
976 SerAspSerThrSerThrSerGlySerValSerGlySerLeuSer..... 990  
3231 ACAGGCACAACTTCGCCCAACACACAGCGGGAAGAACACCGCGCAAA 3280  
991 .....IleAlaAlaSerGlnSerValSerThrSerThrSerAspS 1004  
3281 GCCTT.....CACGCGCTGATTTCGCGCGCGCGC 3309  
1004 erMetSerThrSerGluIleValSerAspSerIleSerThrSerGlySer 1020  
3310 AATCCCAAGGAGCAAGTGTTCGCGAACCGCGCGCGGACGAGG 3359  
1021 LeuSerAlaSerAspSerLysSerMetSerValSerSerSerMetSerTh 1037  
3360 CGGCGAAATTCGCGCATTTATCGAGCGGAGGAGAGACAAAAACGGCTGC 3409  
1037 rSerGlnSerGlySerThrSerGluSerLeuSerAspSerGlnSerThr. 1053  
3410 AGCGGATTAAGACACCGCTTCGCGAAACAGCGCGGAGCGGAAACCGCG 3459  
1054 ..SerAspSerAsp.....SerLysSerLeuSerLeuSerThrSer 1066

## seq\_documentation\_block:

RTX toxin RtxA VC1451 [imported] - Vibrio cholerae (strain M16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: C82199  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
 L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000

A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833

A:Accession: C82199

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4558 <HEI>

A:Cross-references: GB:AE004223; GB:AE003852; NID:g9655942; PIDN:AAF94608.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain M16961; biotype El Tor

C:Genetics:

A:Gene: VC1451

A:Map position: 1

## alignment\_scores:

Quality: 273.50 Length: 1747  
 Ratio: 0.351 Gaps: 82  
 Percent Similarity: 44.591 Percent Identity: 20.034

## alignment\_block:

US-09-303-518D-653 x C82199 ..

Align seg 1/1 to: C82199 from: 1 to: 4558

```

37 CGCAAGCCCTAAACCGCGCCGCTACCG...TTCTCGCCGCTTACTT 83
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365 ArgAsnAspProGlnThrGluLeuLysTyrSerThrAlaTnp.. 380
84 AGCCATATGCTGCTGCTGCGGCTTCTGCCCAAGCCCGCGGACACA 133
380 .....
134 CTTATTTCCGCACTAC.....CAATACTATCGGACTTTGCC 174
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 ..TyrLysGluValAsnHisLeuSerAsnLeuAlaAsnGlnAspLeuSer 396
175 GAAATAAAGCAAGTTTCCAGTC.....GGGGG.....AA 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
397 AspAsnGlyGlyPheThrAlaValAsnHisLeuSerValThrLeuSe 413
207 AGATATGAGTTTACACAAAAAGG.....GAGTTGTCGCA 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
413 rAspLeuLysValGluHisGlnSerValThrValHisAlaValGluL 430
248 AATCGATGACGAAGCCCGATGATGATTTTCTGTGTCGCGTAAC 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 ySerLeuThrGluTyrGluTyrPValThrTyrAla.....Asn 442
298 GCGGTGGCGGATTCGCGGATCAATATATGACGCTGGCACATAA 347
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
443 Gly...AlaValIleAspAlaLysGluValSerLeuSerAlaLysMe 458
348 CGCGCGCTAT.....AACATGTTGATTTTGTGCGG 379
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
458 tGlyGlyHisAlaIleTyrAlaAspGlyThrLysValAspValLysAlaV 475
380 AGGGAAGCAATCCCGATCAGACCGCTTTCTTACCAATTCG..... 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 aLysSerAsnArgGlnProAsnThrTyrIleTyrAlaLysValLeuGly 491
423 .....
492 ProTyrThrLysIleValValValGluLeuAlaAsnAspProGluThrG 508

```

```

424 .....AAAGAAATAATTAAGCAGG..... 447
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
508 yAlaLeuLysTyrGlnAlaArgSerTrpTyrLysGluGlyAspHisThrA 525
448 .....ACTAACGGCCATCTTAT 465
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
525 laAsnIleAlaAsnGlnAspIleSerSerAlaThrGlyTyrAsnPromet 541
466 GGC...GGCGATTATCATATGCGGGTTTGCACAAATTTGTGCACAGATC 512
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
542 GlyLysGlyGlyTyrSerLeuSerAspLeuHisTyrSerValAsnAlaV 558
513 AGAACCTGTTGAGATGACCATTTATATGATGGTGGAAATACGCTGATT 562
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
558 lArgSerThrSerGluThr.....ValAlaAspI 568
563 TAAATAAATACCTCATCTGCTGTCGAATCGGAGCAGCAGCAATATTGG 612
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
568 leGluGluTyrThrAspGlnThrLeuPhe..... 577
613 CGGTCTGATGAAGACGACCAACCAATTAAC.....CGCGAAAGTTTCATATCA 656
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
578 .....LysProAlaAsnAspSerGlyGluSerSerGlyAs 589
657 TATTGCAAGCGCATATCTTGGCTCGTGGCAATACCTTTGCACAAA 706
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
589 pVal.....ArgPheAsnGlyAlaGlyGlyAsnValIleLysSerA 604
707 AT.....GATCAGGTGGTGCACAGTCAACTTA 735
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
604 snValThrArgGlyAsnValHisPheAsnGlyGlyIleAlaAsnVal 620
736 GGTAGCGAAATAATAACAT...AGCCCATATGTTTATACCAACAGG 782
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
621 .....IleLeuHisSerSerGlnPheGly.....As 629
783 AGGCTCATTTGGCGACAGTGGCTCACCATACTTATCTATGATGCCCAA 832
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
629 nThrGluPheAsnGlyGlyAlaAlaAsnValIleValLysSerGlyG 646
833 AGCAAAAGTGGTTAATTAATGGGGTATGCAACAGCAACCCCTATATA 882
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
646 luGluGlyAspLeuThrPheArgGlyAlaGlyLeuAlaAsnValLeuVal 662
883 GGAATAAGCAATGCTTCCAGTAGTTTCGTAAGATTGGTTCTATGATGA 932
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
663 HisGlnSerGluGlnGlyLysMet.....As 671
933 ATCTTTGCTGGAGATACCCATTCAGTATCTACGAACACACATCAAAATG 982
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
671 pValTyrAlaGlyGlyAlaValAsnValLeuValArg...LeuGlyAspG 687
983 GGAATAACTTTTAAAC.....GACAATAATAAT 1011
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
687 lyGlnTyrLeuAlaHisLeuLeuAlaTyrGlyAsnIleSerValGlnLys 703
1012 GGCAGCAAGAAATTCGATGCCAAACATAAACAATACTTCTCTACCTTATAG 1061
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
704 GlySerGly..... 706
1062 ATTAAACACAGAACCGTTCAATTG.....TTTAATGTTTCTTTATCCG 1105
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
707 ....AspSerArgValValMetLeuGlyGlyTyrAsnThrHisThrGlnI 722
1106 AGACAGCAAGAACCTGTTTATCATGCTGCGAGTGGGTCAAC..... 1149
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
722 leGlySerGlyAsnGlyLeuTyrPLeuAlaAlaGlyGlyPheAsnValMet 738
1150 .....AGTTATCGACCCAGACTGAATTAATGGAGAAA 1181
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
739 ThrGlnValGlyLysGlyAspValAlaAlaValLeuAlaGlyGlyAlaAs 755
1182 TATTTCTTTTATTGACAAAGGAAAGGTGAATTGATCTACTTACCAGCA 1231

```

Mon Jul 1 09:27:07 2002

```

1038 aMetPheGlyLysAlaAsnIleMetThrHisValGlyAspGlyLeuThrG 1055
1079 AAATGGAAGGTATCCACAAAGAGAAATCGTGTGGCAACACGATTGGATC 2028
1055 lyValLeuAlaLeuGlyGluAlaAsnIleVal..... 1065
2029 GACCGCACATTTAAAGCGGAAACTTCCATATTCAGCGCGGCAACGCGT 2078
1066 .....ThrLysLeuGlyAspPhe.....MetGlyValAlaAla 1078
2079 GGTTCCTCCGCAATGTTCGCCAAGTGAAGCGCAT..... 2112
1078 aAlaLysAlaAsnValValThrHisValGlyAspAlaThrThrAlaAlaVal 1095
2112 ..... 2112
1095 alLeuAlaGlyLysGlyAsnIleLeuThrLysValGlyGluGlyThrThr 1111
2113 .....TGGCATTTAAGCAATCAGCC..... 2133
1112 ValGlyLeuLeuIleSerAspValGlyAsnValMetThrHisValGlyAs 1128
2134 ..CAAGCAGTTTCGGTCTCCGCCATCAAGCCACACAAATCTGTACAC 2182
1128 pGlyThrIleGlyIleAlaLysGlyLysAlaAsnLeuIleThrLysV 1145
2183 GTTCGGACTGGCGGTCTGACAAAGTTGTACCGGAAAAAACCATTCACGAC 2232
1145 alGlyAspGlyLeuGlyVal.....AsnValThrTrpGlyGln 1157
2233 GATAAAGTGTTCATTGTCAGCAACAGCAGC..... 2265
1158 AlaAsnValPheThrGlnValGlyAspGlyAspArgTyrAsnPheAlaLy 1174
2265 ..... 2265
1174 sGlyGluAlaAsnLeuIleThrLysValGlyAspGlyGlnGluValSerV 1191
2266 ..ATCAGAGCAATGTCAGCTTCGCCATCAGCT..... 2298
1191 alValGlnGlyGluAlaAsnIleIleThrHisValGlyAsnGlyAspAsp 1207
2299 CATTTAAATCTCAGAGACTTGCACACTC..... 2328
1208 TyrThrGlyAlaTrpGlyLysAlaAsnValIleThrLysValGlyHisGl 1224
2328 ..... 2328
1224 yGlnAsnValValLeuAlaLysGlyGluAlaAsnIleValThrGlnValG 1241
2329 .....AACGGCAATCTTAGTGCA 2346
1241 lyAspGlyAspSerPheAsnAlaLeuTrpSerLysGlyAsnIleValThr 1257
2347 GCGGAGACACCCACTATACGTTACGCGCAACGCCACCCCAACGCGCAA 2396
1258 LysValGlyAspGlyMetGlnValThrAlaAlaLysGlyGlnAlaAsnI 1274
2397 CCTCAGCTCGTGGGCAATGCC...CAAGCAACATTTAATCAAGCCACAT 2443
1274 eThrThrThrValGlyAsnGlyLeuAsnValThrAlaAlaTrpGlyAspA 1291
2444 TAAAGCGCAACACATCGGCTTCGGACAATGCTTCATTTAATCTAAGC... 2490
1291 laAsnIleAsnThrLysValGlyAspGlyValSerValAsnValAlaTrp 1307
2491 .....AACACGCGCT 2501
1308 GlyLysTyrAsnIleAsnThrLysValGlyAspGlyLeuAsnValAla 1324
2502 ACAA.....AACGGCAGTCTGACGCTTCCGCAACACGCTAAGG 2539
1324 lMetLysGlyLysAlaAsnAlaAsnIleHisValGlyAspGlyLeuAsnI 1341

```

2540 CAAACGTAAGCCATTCCGCACCTCAAC.....GGCAAT..... 2571  
 ||||:||||:||||: |||  
 1341 leAsnAlaSerTyrAlaGlnAsnAsnValAlaIleIleValGlyAsnGly 1357  
 ||||:||||:||||: |||  
 2572 .....GTCTCCCTAGCC  
 ||||:||||: |||  
 1358 AspPheTyrSerLeuAlaValAlaIleSerSerAsnThrSerSerAsnLysLe 1374  
 ||||:||||: |||  
 2592 AGTATTCCATTGAAACAGCCGCTTTACCGGAAATATCATCCGCGGCCA 2641  
 ||||:||||: |||  
 1374 uSerAlaLeuPheAspAsnIleLysGlnThrValLeuGlyValGlyGly 1391  
 ||||:||||: |||  
 2642 AGGATACGGCATTTACACTTAAAGACAGCAAGTATGACGCTCCCTCGGGC 2691  
 ||||:||||: |||  
 1391 erGlnAlaIleAsnTyrLeuValGlnGlyAspGluAlaSerSerGly 1407  
 ||||:||||: |||  
 2692 ACGGAATTAGGCAATTTAAACCTTGACACGCCACCATTCATCAATTC 2741  
 ||||:||||: |||  
 1408 ThrHisLysGlyArgGlyAlaIleAlaThrProGluIleThrLysLeuAs 1424  
 ||||:||||: |||  
 2742 CGCCTATCGACAGCATGCG.....GCAGCGC 2767  
 ||||:||||: |||  
 1424 pGlyPheGlnMetAspAlaIleLysGluValSerSerAspLeuGlyAspS 1441  
 ||||:||||: |||  
 2768 CGCAAAACCGCAGCTGG.....GCAGATCGCGCGCGCGCTTCGCGC 2811  
 ||||:||||: |||  
 1441 erLeuThrGlySerValThrLysValAspThrProAspLeuAsnLysMet 1457  
 ||||:||||: |||  
 2812 CGTTCCTATTATTCGCTTACGCGCCCACTTCGCGCAATTCGCTTCA 2861  
 ||||:||||: |||  
 1458 GlnHisAlaLeuAsnVal.....AspAspSerValGlnAlaPr 1471  
 ||||:||||: |||  
 2862 CACGCTGACGGTAAACGCCAA.....TTGAAC 2889  
 ||||:||||: |||  
 1471 oAsnLeuIleValAsnGlyAspPheGluLeuGlyHisGlyTrpGlns 1488  
 ||||:||||: |||  
 2884 .....TTGAAC 2889  
 ||||:||||: |||  
 1488 erThrHisGlyValGluAlaSerTyrAlaGlySerValTyrGlyValGlu 1504  
 ||||:||||: |||  
 2890 GGTACAGGAACATTCGCTTTATCTCGCAACTCTTCGGCTACGCGAGCG 2939  
 ||||:||||: |||  
 1505 GlyGluGlyHisGlyAlaArgValThrGluLeuAspThrTyrThrAsnTh 1521  
 ||||:||||: |||  
 2940 CAATATG.....AAGCTGGCGGAAGTTCCGAAGCACTTACACTTGG 2983  
 ||||:||||: |||  
 1521 rSerLeuTyrGlnAspLeuAlaAsnLeuAlaGlnGlyGlu...ValIleA 1537  
 ||||:||||: |||  
 2984 CTGTC.....AACATACCGGC 3000  
 ||||:||||: |||  
 1537 laValSerPheAspPheAlaLysArgAlaGlyLeuSerAsnAsnGluGly 1553  
 ||||:||||: |||  
 3001 AACGAAACCCGTA...AGTCTCGAGCAATTCGAGGTAGTGGAGGAAGAAGA 3047  
 ||||:||||: |||  
 1554 IleGluValLeuTyrAsnGlyGluValValPheSerSerGlyAspGI 1570  
 ||||:||||: |||  
 3048 CAACACACCGCTGCCAAATCTTAATTCACCTCGAAACCAACACAGC 3097  
 ||||:||||: |||  
 1570 uSerAlaTrpGlnGlnLysAsnLeuLysLeuThrAlaGln..... 1583  
 ||||:||||: |||  
 3098 TCGATGCGCGCATGCGCTTATCAGCTTATC.....CGCAAGAC 3138  
 ||||:||||: |||  
 1584 .....AlaGlySerAsnArgIleGluPheLysGlyThrGlyHisAsnAsp 1598  
 ||||:||||: |||  
 3139 GGC..... 3141  
 ||||:||||: |||  
 1599 GlyLeuGlyTyrIleLeuAspAsnValValAlaThrSerGluSerSerGI 1615  
 ||||:||||: |||  
 3142 .....GAGTTCCGCTCATTAATCCGCTCAAGAAC 3172  
 ||||:||||: |||  
 1615 nGlnAlaAsnAlaIleArgGluHisAlaThrGlnAsnProAlaIleGlnA 1632

1903 AlaLeuGluGlyAlaThrAsnAlaValAsnArgLeuGlnIleAsnAlaGI 1919  
 4006 GCAGGTTTCGGCGGATTCGGCATCGAACCGCACATCGCGCAACGGCGCTA 4055  
 1919 yIleArgAlaLysAsnSerValSerSerMetThrSerPheSerGluF 1936  
 4056 TTTCGTCCTCAAAAGCGGATTCACCATACGAAAGCGTCAATATCGGCACCC 4105  
 1936 hrAsnSerLysSerIleValValProThrLysValSerProGluProGlu 1952  
 4106 CGGGCGCTTCATTCACCGCATACCGCGGGCGCATTAAGG 4144  
 1953 Arg.....GlnGluValThrArgArgAspValArg 1962

seq\_name: pir2:B89921

seq\_documentation\_block:  
 hypothetical protein ebha [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: B89921  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; PMID:21311952; PMID:11418146  
 A:Accession: B89921  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-6713 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BAB42527.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: ebha

alignment\_scores:  
 Quality: 273.00 Length: 1667  
 Ratio: 0.351 Gaps: 84  
 Percent Similarity: 46.611 Percent Identity: 20.756

alignment\_block:

US-09-303-518D-653 x B89921 ..

Align seg 1/1 to: B89921 from: 1 to: 6713

193 GCAGTCGGCGGCAAGCATATAGGTTTACACAAAGGAGGAGTTGGT 242  
 1936 hrAsnSerLysSerIleValValProThrLysValSerProGluProGlu 1952  
 306 AlaIleThrGlnAlaThrThrGlnValAsnAsnAlaLysAsnGlyLeuAs 322  
 243 CGGC.....AAATCGATGACGAAAGCCCGATGA 271  
 322 nGlyAlaGluAsnLeuArgAsnAlaGlnAsnThrAlaLysGlnAsnLeuA 339  
 272 TTGATTTTCTGTGATTCGCTACGCGGCGGTGGCGCATTG..... 312  
 339 snThrLeuSerHisLeuThrAsnAsnGlnLysSerAlaIleSerSerGln 355  
 313 .....CGGGCGCATCAATATATTGTGAGCGTGGCGACATAACGGCGG 353  
 356 IleAspArgAlaGlyHisValSerGluValThrAlaAlaLysAsnAlaA 372  
 354 CTATAACANVTTCATTTTGGTGGGAGGAAAGCAATCCGATCAGCACC 403  
 372 a.....ThrGluLeuAsnAlaGlnMetGlyAsnLeuGluGln.... 384  
 404 GCTTTTCTTACCAATTTGTGAAGAAGAAATATATAAGCAGGAGCTAAC 453  
 385 .....AlaIleHisAspGlnAsnThrValLysGlnGlyValAsn 397  
 454 GGCCATCCTTATGGCGCGATATATCATATGCCGCTTGCACAAATTTGT 503

398 .....Ph 398  
 504 CACAGATGCAGAACCTGTTGAGATGACCACTTATATGATGGTGGTGAAT 553  
 398 eThrAspAlaAspLysAlaLysArgAspAlaIleThrAsnAlaValSerA 415  
 554 ACGCTGAT.....TTAATAAATACCTGATCGTGTTCGAATCGGAGCA 597  
 415 rGAlaGluThrIleLeuAsnLysThrGlnGly..... 425  
 598 GGCAGACAATATGGCGGTCTGATGAAGACGACCAATTAACGCGGAAG 647  
 426 .....AlaAsnThrSerLysGlnAspValGluAla 436  
 648 TTCATATCATATTCAGCGCATATTCCTGGCTCGTGGTGGCAAT.... 693  
 436 aIleGlnAsnValThrSerAlaLysAsnAlaLeuAsnGlyAspGlnAsnV 453  
 694 ..ACCTTTTCACAAATGGATCGGTGGTGGCACAGTCAACTTAGGTAGC 741  
 453 aIleThrAsnAlaLysAsnAlaLysAsnAlaLeuAsnAsnLeuThrSer 469  
 742 GAAAAAATTAACATAGCCCATATGTTTTCACACAGGAGCTCAT 791  
 469 ..... 469  
 792 TGGCGACAGTGGCTCACCAGTGTATCTATGATGCCCAAGCAAGT 841  
 470 .....IleAsnAsnAlaGlnLysArgAspL 478  
 842 GGTTA.....ATTAATGGGTA 858  
 478 euThrThrLysIleAspGlnAlaThrThrValAlaGlyValGluAlaVal 494  
 859 TTGCAACACAGC.....AACCCCTATATAGG 884  
 495 SerAsnThrGlyThrGlnLeuAsnThrAlaMetAlaAsnLeuGlnAsnGI 511  
 885 AAAAGCAATGGCTTCCAGCTAGTTCGTAAGATTGGTTCTATGATGAA 934  
 511 yIleAsnAspLysAlaAsnThrLeuAlaSerGluAsnTyrHisAspAla. 527  
 935 TCTTCTCGAGATACCCATTCTACGAAACACACATCAAAATGGG 984  
 528 .....AspSerAspLysLysThrAlaIleThrGlnAlaValThrAsnAla 542  
 985 AAATACTTTTAAACGACAATAAT..... 1008  
 543 GluAsnIleLeuAsnLysAsnSerGlySerAsnLeuAspLysAlaAlaVa 559  
 1009 .....AATGGCGGAGAAATCGATGCCA 1033  
 559 lGluAsnAlaLeuSerGlnValThrAsnAlaLysGlyAlaLeuAsnGlyA 576  
 1034 AACAT..... 1038  
 576 snHisAsnLeuGluGlnAlaLysSerAsnAlaAsnThrThrIleAsnGly 592  
 1039 ...AAACACTATTCTCTACCTTAT.....AGATTAAACACAGAACCT 1079  
 593 LeuGlnHisLeuThrThrAlaGlnLysAspLysLeuLysGlnGlnValGI 609  
 1080 TCAATGTTTAAATGTTTCTTATCCGAGACGACGACGAGAACCTGTTATC 1129  
 609 nGlnAlaGlnAsnValAlaGlyValAspThrValLysSerSerAlaAsnT 626  
 1130 ATGCTGCGAGTGGGTCAACAGTTATCGACCCAGACTG..... 1167  
 626 hrLeuAsnGlyAlaMetGlyThrLeuArgAsnSerIleGlnAspAsnThr 642  
 1168 .....AATATGGAGAAATATTT..... 1185

[illegible]

1905 AACAAACGGCAAACTGTTTTCAGCGGCAGACGACACCGCCAGCTTACA 1954  
 926 ysglnthrGlyGlyAsnThrSerLysAsnValAspAsnAlaLeuAs 942  
 1955 ATCATTT...AGGAACGGGTGTCATAAATGGAAGTATCCCAAGGA 2001  
 942 nAlaValThrArgAlaLysAlaAlaLeuAsnGlyAlaGluAsnLeuArg 959  
 2002 GAAATCGTGTGGGACACACGATTGGATCGACCGCACATTTAAAGCGGAAA 2051  
 959 snAlaLysThrSer..... 963  
 2052 CTTCATATTCAGGCGCGGACAGCGGTGTTCCGCAATGTGCCAAAG 2101  
 964 .....AlaThrAsnThrIleAsnGlyLeuProAsnLeuThrGlnL 977  
 2102 TGGAAAGCGGAT.....TGGCATTTAAGCAATCACGCCCAACGAGTTTC 2145  
 977 euGlnLysAspAsnLeuLysHisGlnValGlnAlaGlnAsnValVal 993  
 2146 GGTGTC.....GCACCGCATCAAGCCACACAACTGTACACGTCGGA 2189  
 994 GlyValAsnGlyValLysAspLysGlyAsnThrLeuAsnThr..... 1007  
 2190 CTGGACGGCTCTGCAACAGTTGTACCGCAAAAAACCATCTACCGAGATAAG 2239  
 1008 .....AlaMetGlyAlaLeuArgThrSerIleGlnAsnAsp.... 1019  
 2240 TGATTGTTTCATTGAGCAAGACCGCATCAGAGGCAATGTTCAGCCTGCC 2289  
 1020 .....AsnThrThrLysThrSerGlnAsnTyrLeuAspAla 1031  
 2290 GATCAGCGTCATTTAAACTCTCAGGACTTCCACACATCAACGGCAATCT 2339  
 1032 SerAspSerAsnLysAsnAsnTyrAsnThrAlaValAsnAsnAlaAsn... 1047  
 2340 TAGTCGACGGGAGACACGCACTATACGGTTCAGCGCACCGCCACCAAA 2389  
 1048 .....GlyValIleAsnAlaThrAsnA 1055  
 2390 ACGCAACCTCAGCCTCGTGGCAATGCC.....CAA 2421  
 1055 snProAsnMetAsp.....AlaAsnAlaIleAsnAspMetAlaAsnGln 1069  
 2422 GCAACATTTAATCAAGCCACATTAACGGCAAC.....ACATCGGC 2462  
 1070 ValAsnThrThrLysAlaAlaLeuAsnGlyAlaGlnAsnLeuAlaGlnAl 1086  
 2463 TTCGGACAAATGCTTCATTTAATCTAAGCAAC.....AACGCGC 2500  
 1086 aLysThrAsnAlaThrAsnThrIleAsnAsnAlaGlnAspLeuAsnGlnL 1103  
 2501 TACAAAAACGGCAGCTG...ACGCTTTCCGACACACGCTAAGGCA..... 2541  
 1103 ysglnLysAspAlaLeuLysThrGlnValAsnAsnAlaGlnArgValSer 1119  
 2542 .....ACGTAAGCCATTCGCA.....CTCAACGCGAATGTCTC 2576  
 1120 AspAlaAsnAsnValGlnHisThrAlaThrGluLeuAsnGlyAlaMetTh 1136  
 2577 CCTAGCCGATAAGCAGCATATCCATTTTGAACACACGCCGCTTTACCGGAA 2626  
 1136 rAlaLeuLysAlaAlaIleAlaAspLysGlyAlaArgThrLysAlaSerGlyA 1153  
 2627 AAATCAGCGCGCGCAAGGATACGGCATTCACATTTAAAGACACGCGAATG 2676  
 1153 snTyrValAsnAlaAspGlnGluLysArgGlnAlaTyrAspSerLysVal 1169  
 2677 ACG.....CTGCGCTCGGCACGGAAATAGGCAATTTAAACCT 2714  
 1170 ThrAsnAlaGluAsnIleLeuAsnGlyThr..... 1179



1700 lThrAsnAla.....LysThrGlnLeuAsnGlyA 1710  
4282 .....AGTGGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4317  
|||::: :::::||::: |||  
1710 snHisAsnLeuGluValAlaLysGlnAsnAlaAsnThrAlaIleAspGly 1726  
4318 TTCACGGCTGCCCTCCACACGGCTGCCGCCCAAGGGGCCGCAA 4359  
:::||| :::::|||  
1727 LeuThr.....SerLeuAsnGlyProGln 1734